

# Gaussian Interference Channel with Intermittent Feedback

Can Karakus, *Student Member, IEEE*, I-Hsiang Wang, *Member, IEEE*, and Suhas Diggavi, *Fellow, IEEE*

**Abstract**—We investigate how to exploit intermittent feedback for interference management by studying the two-user Gaussian interference channel (IC). We approximately characterize (within a universal constant) the capacity region for the Gaussian IC with intermittent feedback. We exactly characterize the capacity region of the linear deterministic version of the problem, which gives us insight into the Gaussian problem. We find that the characterization only depends on the forward channel parameters and the marginal probability distribution of each feedback link. The result shows that passive and unreliable feedback can be harnessed to provide multiplicative capacity gain in Gaussian interference channels. We find that when the feedback links are active with sufficiently large probabilities, the perfect feedback sum-capacity is achieved to within a constant gap. In contrast to other schemes developed for interference channel with feedback, our achievable scheme makes use of *quantize-map-and-forward* to relay the information obtained through feedback, performs forward decoding, and does not use structured codes. We also develop new outer bounds enabling us to obtain the (approximate) characterization of the capacity region.

**Index Terms**—Interference management, interference channel, intermittent feedback, unreliable feedback, quantize-map-and-forward

## I. INTRODUCTION

The simplest information theoretic model for studying interference is the two-user Gaussian *interference channel* (IC). It has been shown that feedback can provide an unbounded gain in capacity for two-user Gaussian interference channels [1], in contrast to point-to-point memoryless channels, where feedback gives no capacity gain [2], and multiple-access channels, where feedback can at most provide *power* gain [3]. This has been demonstrated when the feedback is unlimited, perfect, and free of cost in [1]. Given the optimistic result obtained under this setting, a natural question arises: Can feedback be leveraged for interference management under imperfect feedback models?

There have been several pieces of work so far, attempting to answer this question. Vahid *et al.* [4] considered a rate-limited feedback model, where the feedback links are modeled

as fixed-capacity deterministic bit pipes. They developed a scheme based on decode-and-forward at transmitters and lattice coding to extract the helping information in the feedback links, and showed that it achieves the sum-capacity to within a constant gap. The work in [5] studied a deterministic model motivated by passive feedback over AWGN channels, and [6], [7] studied the two-way interference channel, where the feedback is provided through a backward interference channel that occupies the same resource as the forward channel. [5], [6] and [7] only dealt with the linear deterministic model [8] of the Gaussian IC.

In this paper, we investigate how to exploit *intermittent* feedback for managing interference [14], [15]. Such intermittent feedback could occur in several situations. For example, one could use a side-channel such as WiFi for feedback; in this case since the WiFi channel is best-effort, dropped packets might cause intermittent feedback. In other situations, control mechanisms in higher network layers could cause the feedback resource to be available intermittently. For the feedback links, Bernoulli processes  $\{S_1[t]\}$  and  $\{S_2[t]\}$  control the presence of feedback for user 1 and 2, respectively. The two processes can be dependent, but their joint distribution is i.i.d. over time. We assume that the receivers are *passive*: they simply feed back their received signals to the transmitters without any processing. In other words, each transmitter receives from feedback an observation of the channel output of its own receiver through an erasure channel, with unit delay. We focus on the passive feedback model as the intermittence of feedback is motivated by the availability of feedback resources (either through use of best-effort WiFi for feedback or through feedback resource scheduling). Therefore, it might be that the time-variant statistics of the intermittent feedback are not *a priori* available at the receiver, precluding active coding<sup>1</sup>. Moreover, the availability of the feedback resource may not be known ahead of transmission, therefore motivating the assumption of causal state information at the transmitter. If the receiver has *a priori* information about the feedback channel statistics, it can perform active coding, in which case, the intermittent feedback model reduces to the rate-limited model

The research of C. Karakus and S. Diggavi was supported in part by NSF grant 1314937 and a gift from Intel. The work of I.-H. Wang was supported by Ministry of Science and Technology, Taiwan, under Grants MOST 103-2221-E-002-089-MY2 and MOST 103-2622-E-002-034. This paper was presented in part at 2013 IEEE International Symposium on Information Theory and 2013 Allerton Conference on Communication, Control, and Computing.

C. Karakus and S. Diggavi are with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095 USA (e-mail: karakus@ucla.edu; suhas@ee.ucla.edu)

I.-H. Wang is with the Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan (e-mail: ihwang@ntu.edu.tw)

Copyright (c) 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

<sup>1</sup>In general, the statistics of  $S_1[t]$  and  $S_2[t]$  can have arbitrary time-dependence, which could be unavailable at the receivers before feedback transmission, but this information could be learned after the transmission. This implies that receivers may not be able to actively code the feedback signals, but the feedback statistics can potentially be used at the transmitters and later on at the receivers, after the feedback transmissions (therefore the transmitters could use these statistics to encode after receiving feedback). In this work, we focus on the case where feedback statistics is time-invariant for simplicity, but the schemes described here can be easily extended into the time-variant case.

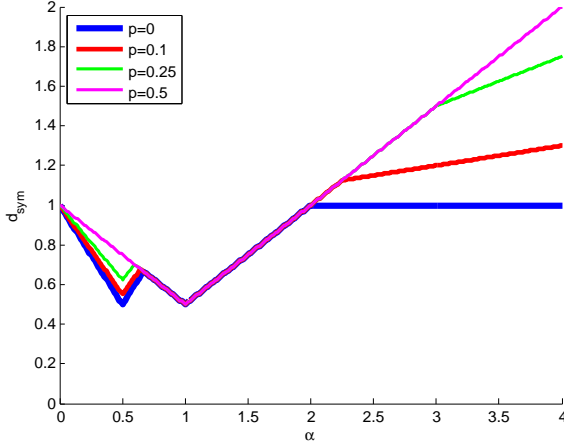


Fig. 1. Generalized degrees of freedom per user with respect to interference strength  $\alpha := \frac{\log \text{INR}}{\log \text{SNR}}$  for symmetric channel parameters.

of [4].

We study the effect of intermittent feedback for the two-user Gaussian IC inspired by ideas we develop for the linear deterministic IC model [8]. Our main contribution is the approximate characterization of the capacity region of the interference channel with intermittent feedback, under the Gaussian model. We also derive an exact characterization of the capacity region under the linear deterministic model, which agrees with the Gaussian result. The capacity characterizations under both models depend only on the forward channel parameters and the marginal distributions of  $S_1$  and  $S_2$ ; not on their joint distribution.

Our result shows that feedback can be harnessed to provide multiplicative gain in Gaussian interference channel capacity even when it is unreliable and intermittent. The result can be interpreted using the picture given in Figure 1, which is depicted (for convenience) in terms of symmetric generalized degrees of freedom for the special case of symmetric channel parameters. The given GDoF curves suggest that as the feedback probability increases, the achievable GDoF also increases for all interference regimes for which perfect feedback provides any GDoF gain. One can also observe from the figure that the capacity gain from intermittent feedback, which depends on the portion of time when the feedback is active, remains unbounded, similar to the perfect feedback case.

A consequence of this result is that when the feedback links are active with large enough probabilities, the sum-capacity of the perfect feedback channel can be achieved to within a constant gap. Similarly for the linear deterministic case, the perfect feedback capacity is exactly achieved even when there is only intermittent feedback, with large enough “on” probability. In particular, under the symmetric setting, this threshold is  $1/2$  for each feedback link. This is also reflected in Figure 1, where the “V-curve” achievable with perfect feedback is already achievable when the feedback probability is only  $1/2$ .

Our achievable scheme has three main differences from the

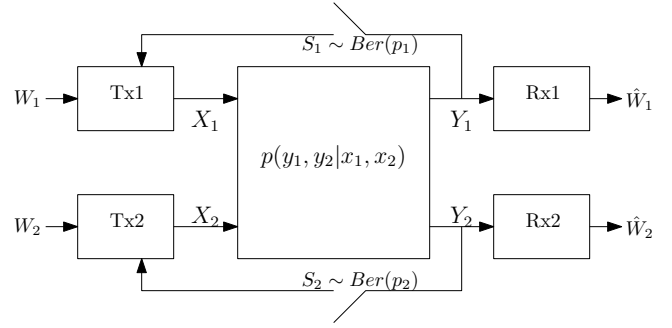


Fig. 2. Two-user discrete memoryless interference channel with intermittent feedback

previous schemes developed in [1], [4] and [5]. First, we use *quantize-map-and-forward* (QMF)<sup>2</sup> [8] at the transmitters to send the information obtained through feedback, as opposed to (partial or complete) decode-and-forward, which has been used in [1], [4], [5]. This is because when there is intermittent feedback, the transmitters might not be able to decode the other user’s (partial) message, but would still need to send useful information about the interference. A similar situation arises in a relay network, where QMF enables forwarding of evidence, without requiring decoding [8]. Second, at the receivers, we perform forward decoding of blocks instead of backward decoding, which results in a better delay performance. Third, we do not use structured codes, *i.e.*, we only perform random coding.

We also develop novel outer bounds that are within a constant of the achievable rate region for the Gaussian IC and match the achievable region for the linear deterministic IC. These outer bounds are based on constructing an enhanced channel and appropriate side-information. These are illustrated in Section VI.

Lastly, we extend these results for packet transmission channels, modeled through parallel channels which are  $M$ -symbol extensions of the original model. This can be considered as a model for OFDM and packet drops over a best-effort channel.

The rest of this paper is organized as follows. We formally state the problem and establish the notation in Section II. We present our main results in Section III and give interpretations of them. We motivate our coding scheme and explain it through an example in Section IV. We give the analysis of the coding scheme in Section V. The outer bound is developed in Section VI and Section VII concludes the paper with a brief discussion of possible extensions of the work. Many of the detailed proofs are given in the Appendices.

## II. SYSTEM MODEL

We consider the 2-user discrete memoryless interference channel (DM-IC) with intermittent feedback, illustrated in Figure 2. We assume Transmitter  $i$  ( $\text{Tx}i$ ) has a message  $W_i$  intended for Receiver  $i$  ( $\text{Rx}i$ ),  $i = 1, 2$ .  $W_1 \in [2^{NR_1}]$  and

<sup>2</sup>The QMF scheme of [8] was generalized to DMCs in [9] (and the scheme was called noisy network coding) and to lattices in [10], [11]. In this paper we use the “short-messaging” version of QMF [12] instead of the “long-messaging” version first studied in [8] and extended to DMCs in [9]. For a longer discussion about this and other issues, refer to Section VII.

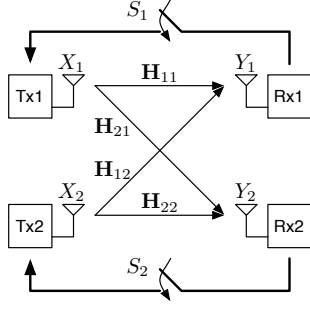


Fig. 3. Two-user linear deterministic interference channel with intermittent feedback

$W_2 \in [2^{NR_2}]$  are independent and uniformly distributed, where, for  $n \in \mathbb{N}$ ,  $[n] := \{k \in \mathbb{N} : k \leq n\}$ . The signal transmitted by Tx $i$  at time  $t$  is denoted by  $X_{i,t} \in \mathcal{X}_i$ , while the channel output observed at Rx $i$  is denoted by  $Y_{i,t} \in \mathcal{Y}_i$ , for  $i = 1, 2$ . For a block length  $N$ , the conditional probability distribution mapping the input codeword to the output sequence is given by

$$p(Y_1^N, Y_2^N | X_1^N, X_2^N) = \prod_{t=1}^N p(Y_{1,t}, Y_{2,t} | X_{1,t}, X_{2,t})$$

The feedback state sequence pair  $\underline{S} := (S_1^N, S_2^N)$  have the joint distribution

$$p(S_1^N, S_2^N) = \prod_{t=1}^N p(S_{1,t}, S_{2,t}).$$

and marginally, at time  $t$ ,  $S_{i,t} \sim \text{Bernoulli}(p_i)$ , for  $i = 1, 2$ , for all  $t$  and  $N$ . Note that, for any fixed time slot  $t$ , the random variables  $S_{1,t}$  and  $S_{2,t}$  are not necessarily independent, that is, the joint distribution  $p(S_{1,t}, S_{2,t})$  can be arbitrary. We assume that receivers have access to  $\underline{S}$  strictly causally, that is, at time  $t$ , both receivers know the realization of  $\underline{S}^{t-1}$ .

At the beginning of time  $t$ , Tx $i$  observes the channel output received by Rx $i$  at time  $t-1$  through an erasure channel, i.e., it receives  $\tilde{Y}_{i,t-1} := S_{i,t-1}Y_{i,t-1}$ , for  $i = 1, 2$ . Note that this is a *passive* feedback model, in that it does not allow the receiver to perform any processing on the channel output; it simply forwards the received signal  $Y_i$  at every time slot, which gets erased with probability  $1 - p_i$ .

For random variables  $A$  and  $B$ , we use the notation  $A \stackrel{f}{=} B$  to denote that  $A$  is a deterministic function of  $B$ <sup>3</sup>. Then our channel model implies  $X_{i,t} \stackrel{f}{=} (W_i, S_i^{t-1}, \tilde{Y}_i^{t-1})$ .

A rate pair  $(R_1, R_2)$  is said to be achievable if there exists a pair of codebooks  $(\mathcal{C}_1, \mathcal{C}_2)$  at Tx1 and Tx2, with rates  $R_1$  and  $R_2$ , respectively, and pairs of encoding and decoding functions such that the average probability of error at any decoder goes to zero as the block length  $N$  goes to infinity. The capacity region with feedback probabilities  $p_1$  and  $p_2$ ,  $\mathcal{C}(p_1, p_2)$ , is defined as the closure of the set of all

<sup>3</sup>More formally,  $A \stackrel{f}{=} B$  means that there exists a  $\sigma(B)$ -measurable function  $f$  such that  $A = f(B)$  almost surely, where  $\sigma(B)$  is the sigma-algebra generated by  $B$ .

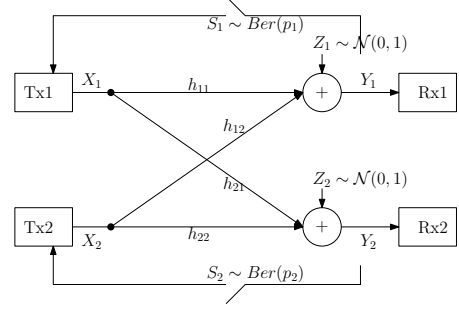


Fig. 4. Two-user Gaussian interference channel with intermittent feedback

achievable rate pairs  $(R_1, R_2)$  when  $S_1 \sim \text{Bernoulli}(p_1)$  and  $S_2 \sim \text{Bernoulli}(p_2)$ . Sum-capacity is defined by

$$C^{\text{sum}}(p_1, p_2) := \sup \{R_1 + R_2 : (R_1, R_2) \in \mathcal{C}(p_1, p_2)\}.$$

In this work, we consider two specific channel models (that is, two specific classes of  $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_1, \mathcal{Y}_2, p(y_1, y_2 | x_1, x_2))$ ), described in the following subsections.

#### A. Linear Deterministic Model

This channel model was introduced in [8] and since then proved useful in providing insight into the nature of signal interactions many network information theory problems (see Figure 3).

We assume  $X_{i,t} \in \mathbb{F}_2^q$ , for  $i = 1, 2$ , where  $\mathbb{F}_2$  is the binary field. The received signal at Rx $i$  is given by

$$Y_{i,t} = \mathbf{H}_{ii}X_{i,t} + \mathbf{H}_{ij}X_{j,t}$$

for  $(i, j) = (1, 2), (2, 1)$ . The channel matrices are given by  $\mathbf{H}_{ij} := \mathbf{S}^{q-n_{ij}}$  for  $(i, j) \in \{1, 2\}^2$ , where  $q = \max\{n_{11}, n_{12}, n_{21}, n_{22}\}$ , and  $\mathbf{S} \in \mathbb{F}_2^{q \times q}$  is the shift matrix  $\begin{bmatrix} \mathbf{0}^T & \mathbf{0} \\ \mathbf{I}_{q-1} & \mathbf{0} \end{bmatrix}$ , where  $\mathbf{0}$  is the zero vector in  $\mathbb{F}_2^{q-1}$  and  $\mathbf{I}_{q-1}$  is the identity matrix in  $\mathbb{F}_2^{(q-1) \times (q-1)}$ . We also define, for  $(i, j) = (1, 2), (2, 1)$ ,

$$V_{i,t} = \mathbf{H}_{ji}X_{i,t}.$$

The capacity region for the linear deterministic model will be denoted by  $\mathcal{C}_{LDC}(p_1, p_2)$ , while its sum-capacity will be denoted by  $C_{LDC}^{\text{sum}}(p_1, p_2)$ .

#### B. Gaussian Model

Under the canonical Gaussian model (see Figure 4), the channel outputs are related to the inputs through the equations

$$\begin{aligned} Y_{1,t} &= h_{11}X_{1,t} + h_{12}X_{2,t} + Z_{1,t} \\ Y_{2,t} &= h_{21}X_{1,t} + h_{22}X_{2,t} + Z_{2,t} \end{aligned}$$

where  $h_{ij} \in \mathbb{C}$ , for  $(i, j) \in \{1, 2\}^2$ , are channel gains, and  $Z_{1,t}, Z_{2,t} \sim \mathcal{CN}(0, 1)$  are circularly symmetric complex white Gaussian noise. We assume an average transmit power constraint of  $P_i$  at Tx $i$ , i.e., for any length- $N$  codeword  $X_i^N$  transmitted by Tx $i$ ,  $\frac{1}{N} \sum_{t=1}^N |X_{i,t}|^2 \leq P_i$ ,  $i = 1, 2$ . We also define

$$\text{SNR}_i := |h_{ii}|^2 P_i$$

$$R_1 \leq \min \{ \max(n_{11}, n_{12}), n_{11} + p_2(n_{21} - n_{11})^+ \} \quad (1)$$

$$R_2 \leq \min \{ \max(n_{22}, n_{21}), n_{22} + p_1(n_{12} - n_{22})^+ \} \quad (2)$$

$$R_1 + R_2 \leq \min \{ \max(n_{11}, n_{12}) + (n_{22} - n_{12})^+, \max(n_{22}, n_{21}) + (n_{11} - n_{21})^+ \} \quad (3)$$

$$R_1 + R_2 \leq \max \{ n_{12}, (n_{11} - n_{21})^+ \} + \max \{ n_{21}, (n_{22} - n_{12})^+ \} \\ + p_1 \min \{ n_{12}, (n_{11} - n_{21})^+ \} + p_2 \min \{ n_{21}, (n_{22} - n_{12})^+ \} \quad (4)$$

$$2R_1 + R_2 \leq \max(n_{11}, n_{12}) + \max \{ n_{21}, (n_{22} - n_{12})^+ \} + (n_{11} - n_{21})^+ + p_2 \min \{ n_{21}, (n_{22} - n_{12})^+ \} \quad (5)$$

$$R_1 + 2R_2 \leq \max(n_{22}, n_{21}) + \max \{ n_{12}, (n_{11} - n_{21})^+ \} + (n_{22} - n_{12})^+ + p_1 \min \{ n_{12}, (n_{11} - n_{21})^+ \} \quad (6)$$

$$\text{INR}_i := |h_{ij}|^2 P_j$$

and

$$V_i := h_{ji} X_i + Z_j,$$

$$\tilde{V}_i := S_j V_i,$$

for  $(i, j) = (1, 2), (2, 1)$ . Note that this definition of  $V_{i,t}$  is consistent with its definition under linear deterministic model, in the sense that it is what remains out of the channel output when the intended signal is completely cancelled.

The capacity region for the Gaussian model will be denoted by  $\mathcal{C}_G(p_1, p_2)$ , while its sum-capacity will be denoted by  $C_G^{\text{sum}}(p_1, p_2)$ . We will also use the notation  $C_{G,p}^{\text{sum}} := C_G^{\text{sum}}(1, 1)$ , denoting the sum-capacity under perfect feedback.

Gaussian parallel channel is described by the equations

$$\mathbf{Y}_{1,t} = h_{11}\mathbf{X}_{1,t} + h_{12}\mathbf{X}_{2,t} + \mathbf{Z}_{1,t} \quad (7)$$

$$\mathbf{Y}_{2,t} = h_{21}\mathbf{X}_{2,t} + h_{22}\mathbf{X}_{1,t} + \mathbf{Z}_{2,t} \quad (8)$$

$$\tilde{\mathbf{Y}}_{1,t} = S_{1,t}\mathbf{Y}_{1,t} \quad (9)$$

$$\tilde{\mathbf{Y}}_{2,t} = S_{2,t}\mathbf{Y}_{2,t} \quad (10)$$

where  $\mathbf{X}_{i,t}, \mathbf{Y}_{i,t} \in \mathbb{C}^M$ ,  $i = 1, 2$ , are the channel input and output, respectively, at user  $i$ ;  $\mathbf{Z}_{1,t}$  and  $\mathbf{Z}_{2,t}$  are independent and distributed with  $\mathcal{CN}(\mathbf{0}, \mathbf{I})$ ; and  $\tilde{\mathbf{Y}}_{i,t}, i = 1, 2$  is the output of the feedback channel of Tx $i$ , at time  $t$ . Note that the channel gains are scalars. It should also be noted that any given time, the same feedback state variable  $S_{i,t}$  controls the presence of feedback for all sub-channels, *i.e.*, the feedback is present either for all  $M$  channels, or for none of them.

### III. MAIN RESULTS

In this section, we present our results and discuss their consequences for both linear deterministic and Gaussian models.

#### A. Linear Deterministic Model

The following theorem captures our main result for the linear deterministic model.

**Theorem 3.1.** *The capacity region  $\mathcal{C}_{LDC}(p_1, p_2)$  of the linear deterministic interference channel with intermittent feedback is given by the set of rate pairs  $(R_1, R_2)$  satisfying (1)–(6).*

*Proof.* See Section V for achievability, and Section VI for converse.  $\square$

Note that for the special cases of  $p_1 = p_2 = 1$  and  $p_1 = 1, p_2 = 0$ , existing results in the literature [1], [6] are recovered. The following corollary shows that it is possible to achieve perfect feedback sum-capacity even when feedback probabilities are less than one.

**Corollary 3.1.** *For  $n_{12}, n_{21} > 0$ , there exists  $p^* < 1$  such that*

$$C_{LDC}^{\text{sum}}(p_1, p_2) = C_{LDC}^{\text{sum}}(1, 1)$$

for all  $p_1, p_2 \geq p^*$ .

*Proof.* See Appendix G.  $\square$

We illustrate Corollary 3.1 through an example. Let us assume  $n_{12} = n_{21} = m$ ,  $n_{11} = n_{22} = n$ , and  $p_1 = p_2 = p$ . It is easy to see that if  $p_1 = p_2 = 0.5$ , the bounds on  $R_1 + R_2$ ,  $2R_1 + R_2$  and  $R_1 + 2R_2$  that involve  $p_1$  and  $p_2$  become redundant, and the sum-capacity does not increase beyond this point, for all  $(m, n)$ .

#### B. Gaussian Model

We define, for any set  $\mathcal{R}$  of rate pairs  $(R_1, R_2)$  and scalar  $\delta \in \mathbb{R}$ ,

$$\mathcal{R} - \delta := \{ (R_1, R_2) : (R_1 + \delta, R_2 + \delta) \in \mathcal{R} \},$$

$$\mathcal{R} + \delta := \{ (R_1, R_2) : (R_1 - \delta, R_2 - \delta) \in \mathcal{R} \}.$$

The following theorem captures our main result for the Gaussian model.

**Theorem 3.2.** *The capacity region  $\mathcal{C}_G(p_1, p_2)$  of the Gaussian interference channel with intermittent feedback satisfies*

$$\bar{\mathcal{C}}(p_1, p_2) - \delta_1 \subseteq \mathcal{C}_G(p_1, p_2) \subseteq \bar{\mathcal{C}}(p_1, p_2) + \delta_2 \quad (15)$$

where  $\bar{\mathcal{C}}(p_1, p_2)$  is the set of rate pairs satisfying (11)–(14) for  $(i, j) = (1, 2), (2, 1)$  and  $\delta_1 < 2 \log 3 + 3(p_1 + p_2)$  bits, and  $\delta_2 < \log 3 + p_1 + p_2$  bits.

*Proof.* Section V proves an inner bound region  $\mathcal{R}_G^i(p_1, p_2)$ , Section VI proves an outer bound region  $\mathcal{R}_G^o(p_1, p_2)$ , and Appendix F shows that  $\bar{\mathcal{C}}(p_1, p_2) - \delta_1 \subseteq \mathcal{R}_G^i(p_1, p_2)$  and  $\mathcal{R}_G^o(p_1, p_2) - \delta_2 \subseteq \bar{\mathcal{C}}(p_1, p_2)$ .  $\square$

**Remark 3.1.** *Theorem 3.2 uniformly approximates the capacity region under Gaussian model to within a gap of  $3 \log 3 + 4(p_1 + p_2)$  bits, independent of channel parameters. To our knowledge, this is the first constant-gap capacity region*

$$R_i < \min \left\{ \log(1 + \text{SNR}_i + \text{INR}_i), \log(1 + \text{SNR}_i) + p_j \log \left( 1 + \frac{\text{INR}_j}{1 + \text{SNR}_i} \right) \right\} \quad (11)$$

$$R_i + R_j < \log \left( 1 + \frac{\text{SNR}_i}{1 + \text{INR}_j} \right) + \log(1 + \text{SNR}_j + \text{INR}_j) \quad (12)$$

$$R_i + R_j < \log \left( 1 + \frac{\text{SNR}_i}{1 + \text{INR}_j} + \text{INR}_i \right) + \log \left( 1 + \frac{\text{SNR}_i}{1 + \text{INR}_j} + \text{INR}_i \right) \\ + p_i \log \left( \frac{(1 + \text{INR}_i) \left( 1 + \frac{\text{SNR}_i}{1 + \text{INR}_j} \right)}{1 + \frac{\text{SNR}_i}{1 + \text{INR}_j} + \text{INR}_i} \right) + p_j \log \left( \frac{(1 + \text{INR}_j) \left( 1 + \frac{\text{SNR}_j}{1 + \text{INR}_i} \right)}{1 + \frac{\text{SNR}_j}{1 + \text{INR}_i} + \text{INR}_j} \right) \quad (13)$$

$$2R_i + R_j < \log \left( 1 + \frac{\text{SNR}_i}{1 + \text{INR}_j} \right) + \log \left( 1 + \frac{\text{SNR}_j}{1 + \text{INR}_i} + \text{INR}_j \right) \\ + \log(1 + \text{SNR}_i + \text{INR}_i) + p_j \log \left( \frac{(1 + \text{INR}_j) \left( 1 + \frac{\text{SNR}_j}{1 + \text{INR}_i} \right)}{1 + \frac{\text{SNR}_j}{1 + \text{INR}_i} + \text{INR}_j} \right) \quad (14)$$

characterization for Gaussian interference channel with non-perfect feedback with arbitrary channel parameters.

**Remark 3.2.** As will be seen in the achievability proof, the proposed coding scheme achieves a smaller gap than what is given in Theorem 3.2; however, for simplicity in the achievability proof, we lower bound the achievable rate terms with computationally more tractable ones, which artificially contributes to the claimed gap. Moreover, one can optimize over the parameters of the proposed coding scheme, such as power allocation and quantization distortion, to further reduce the gap, but this issue will not be dealt with in this paper.

Theorem 3.2 allows us to characterize the symmetric generalized degrees of freedom under symmetric channel parameters, which is a metric often used to compare the capabilities of the interference channel under different settings.

**Corollary 3.2** (Generalized Degrees of Freedom). *For symmetric channel parameters ( $\text{SNR}_1 = \text{SNR}_2 = \text{SNR}$ ,  $\text{INR}_1 = \text{INR}_2 = \text{INR}$ ,  $p_1 = p_2 = p$ ), the symmetric generalized degrees of freedom of freedom, defined by*

$$d_{\text{sym}} := \lim_{\substack{\text{SNR} \rightarrow \infty \\ \text{INR} = \text{SNR}^\alpha}} \frac{C_{\text{sym}}(\text{SNR}, \text{INR}, p)}{\log \text{SNR}},$$

where  $C_{\text{sym}}(\text{SNR}, \text{INR}, p) := \sup \{R : (R, R) \in \mathcal{C}_G(p, p)\}$ , is given by

$$d_{\text{sym}} = \begin{cases} \min\{1 - \alpha/2, 1 - (1 - p)\alpha\}, & \alpha \leq 1/2 \\ \min\{1 - \alpha/2, p + (1 - p)\alpha\}, & 1/2 \leq \alpha \leq 1 \\ \min\{\alpha/2, (1 - p) + p\alpha\}, & \alpha \geq 1 \end{cases}$$

Figure 1 plots the available generalized degrees of freedom with respect to interference strength for various values of  $p$ . As can be observed, as  $p$  is increased, gradually better curves are obtained. It should be noted that once  $p \geq 0.5$ , the “V-curve” that is achieved by perfect feedback [1] is already achieved. Next, this observation will be made precise.

The perfect feedback outer bound on the sum-capacity,  $C_{G,p}^{\text{sum}}$ , is given in Theorem 3 of [1] as follows.

$$C_{G,p}^{\text{sum}} \leq \sup_{0 \leq \rho \leq 1} \min_{(i,j) \in \mathcal{I}} \log \left( 1 + \frac{(1 - \rho^2)\text{SNR}_i}{1 + (1 - \rho^2)\text{INR}_j} \right)$$

$$+ \log \left( 1 + \text{SNR}_j + \text{INR}_j + 2\rho\sqrt{\text{SNR}_j \cdot \text{INR}_j} \right)$$

where  $\mathcal{I} = \{(1, 2), (2, 1)\}$ . The next corollary shows that when  $p_1$  and  $p_2$  are sufficiently large, the sum-capacity of the perfect feedback Gaussian channel can be achieved with intermittent feedback, to within a constant gap. Hence, this corollary is the Gaussian counterpart of the similar result given in Corollary 3.1, for the linear deterministic channel.

**Corollary 3.3.** *For  $\text{INR}_1, \text{INR}_2 > 0$ , there exists  $p^* < 1$  such that*

$$C_{G,p}^{\text{sum}} - C_G^{\text{sum}}(p_1, p_2) \leq \delta_p$$

for all  $p_1, p_2 \geq p^*$ , where  $\delta_p$  is a constant independent of channel parameters.

*Proof.* See Appendix G.  $\square$

In our intermittent feedback model, erasures are symbol-wise, that is, each symbol can get erased independently of others. However, in a best-effort channel, erasures might occur on packet-level instead. In order to study this scenario, we consider the parallel channel model described by the equations (7)–(10), which is simply the  $M$ -symbol extension of the Gaussian channel, where the channel parameters are the same for each subchannel. Each extended symbol over this channel models a packet. The result in Theorem 3.2 easily generalizes to parallel channel model, as shown by the following corollary.

**Corollary 3.4** (Parallel channel). *The capacity region  $\mathcal{C}_G^{(M)}(p_1, p_2)$  of any parallel channel of size  $M$  with feedback probabilities  $p_1$  and  $p_2$  satisfies*

$$M\bar{\mathcal{C}}(p_1, p_2) - M\delta_1 \subseteq \mathcal{C}_G^{(M)}(p_1, p_2) \subseteq M\bar{\mathcal{C}}(p_1, p_2) + M\delta_2$$

where  $\bar{\mathcal{C}}(p_1, p_2)$ ,  $\delta_1$  and  $\delta_2$  are as defined in Theorem 3.2.

**Remark 3.3.** Although strictly speaking, the claim in Corollary 3.4 is more general than that in Theorem 3.2, the achievability and converse proofs for the scalar channel directly extend to the parallel channel without any non-trivial modification. Hence, for simplicity, we focus on the scalar

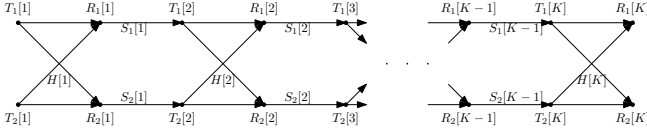


Fig. 5. The interference network unfolded over a block of  $K$  time slots. The node  $T_i[t]$  corresponds to the copy of  $T_{xi}$  at time  $t$ , while  $R_i[t]$  corresponds to the copy of  $R_{xi}$  at time  $t$ . The feedback channel for time  $t$  is an erasure channel controlled by  $S_1[t]$  and  $S_2[t]$ , while the forward channel is a Gaussian interference channel with channel matrix  $H[t]$ .

case in the paper, and omit a separate proof for the parallel channel.

### C. Discussion of Results

1) *Feedback Strategy*: Our result shows that even unreliable feedback provides multiplicative gain in interference channels. The key insight in showing this result is using quantize-map-forward as a feedback strategy at the transmitters. This is in contrast to the schemes proposed for perfect feedback [1] and rate-limited feedback [4], which use decode-and-forward to extract the feedback information. When the feedback channel is noisy<sup>4</sup>, such schemes can result in rates arbitrarily far from optimality. In order to see this, consider unfolding the channel over time, as shown in Figure 5. This transformation effectively turns this channel into a relay network, where it is known that decode-and-forward based relaying schemes can give arbitrarily loose rates. This also motivates using quantize-map-forward as a feedback strategy, which has been shown to approximately achieve the relay network capacity [8]. This observation also suggests that quantize-map-forward might be a promising feedback strategy for the additive white Gaussian noise (AWGN) feedback model of [5] in order to uniformly achieve its capacity region to within a constant gap.

It is instructive to compare the achievable rate region for the case of  $p_1 = p_2 = 1$  with the outer bound region of the perfect feedback model of [1]. Evaluating the region  $\bar{C}(p_1, p_2) - \delta_1$  with  $p_1 = p_2 = 1$ , we see that the perfect feedback bound (12) becomes redundant, and the achievable region comes within  $(3 + 3 \log 3)$  bits of the outer bound region of [1] (see Appendix F for details). We note that this gap is larger than what is achieved by the decode-and-forward based scheme of [1]. This shows that uniform approximation of capacity region via quantize-map-forward comes at the expense of an additional (but constant) gap<sup>5</sup>. The source of this additional gap is the quantization step at the transmitters, which introduces a distortion in the feedback signal, and eventually incurs a constant rate penalty whose amount depends on the distortion level.

2) *Perfect Feedback Sum Capacity with Intermittent Feedback*: Corollary 3.3 shows that for any set of channel parameters, there exists a threshold  $p^*$  on the feedback probability above which perfect feedback sum-capacity is achieved to

<sup>4</sup>“Noise” in this context refers to the erasures in the feedback channel.

<sup>5</sup>Although we stated that the quantize-map-forward scheme achieves a smaller gap than what is claimed in Theorem 3.2, the actual gap is still expected to be larger than that of the decode-and-forward based scheme for perfect feedback, due to quantization distortion.

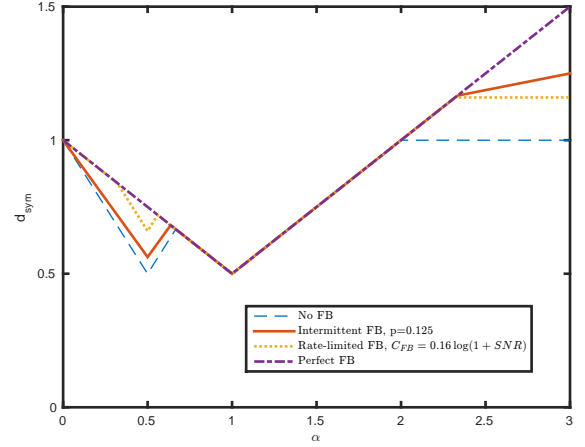


Fig. 6. Generalized degrees of freedom per user with respect to interference strength  $\alpha := \frac{\log \text{INR}}{\log \text{SNR}}$  for symmetric channel parameters, for no feedback, intermittent feedback, rate-limited feedback and perfect feedback.

within a constant gap. Although the exact closed-form expression of  $p^*$  is not clean, an examination of the symmetric case (see Figure 6) reveals that in some cases it can be as low as 0.5.

The intuition behind this result lies in the fact that it takes the transmitter forward-channel resources to send the information obtained through feedback. Note that the larger  $p$  is, the larger the amount of additional information about the past reception can be obtained through intermittent feedback at the transmitters. If the amount of such information is larger than a threshold, then sending it to the receivers will limit the rate for delivering fresh information. Hence, once this threshold is reached, having more feedback resource is no longer useful. However, this property is not observed for the entire capacity region, since if one of the users transmit at a low rate, then it will have sufficient slackness in rate to forward the entire feedback information.

3) *Comparison with Rate-Limited Feedback*: Given that both intermittent feedback and rate-limited feedback of [4] provide degrees-of-freedom gains, one might ask how the two models compare. In order to understand the relative merits of the two feedback models, we revisit the symmetric generalized-degrees-of-freedom curves for the two models for symmetric channel parameters, plotted in Figure 6. The figure illustrates the fact that there is no direct equivalence between the two models, *i.e.*, there is no amount of rate-limited feedback that can exactly replicate the gain of intermittent feedback uniformly for all interference strengths, and vice versa. For the specific feedback parameters given in Figure 6, we observe that rate-limited feedback is more useful for the weak interference regime described by  $\alpha \leq 2/3$ , whereas intermittent feedback is more useful for the strong interference regime, given by  $\alpha \geq 2$ . The reason is that for weak interference, rate-limited model allows for block processing of the channel output to generate the feedback signal, hence feedback helps to resolve the interference in all time slots, whereas for intermittent feedback, information about interference in some time slots gets unrecoverably lost on the erasure channel. On

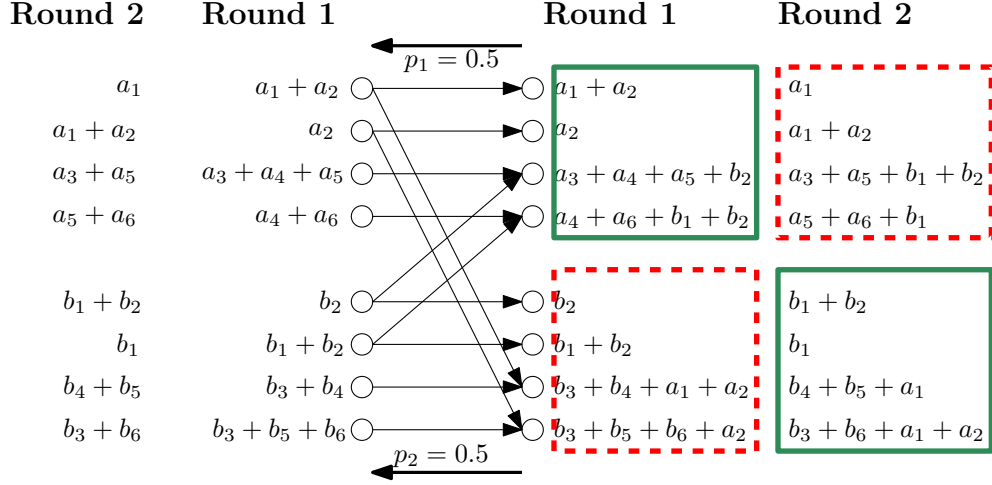


Fig. 7. First block of transmissions for the example coding scheme over linear deterministic channel. Receptions enclosed in green/solid rectangles represent the channel outputs that the receivers are able to feed back; whereas those enclosed in red/dashed rectangles represent the channel outputs that gets erased through the feedback channel.

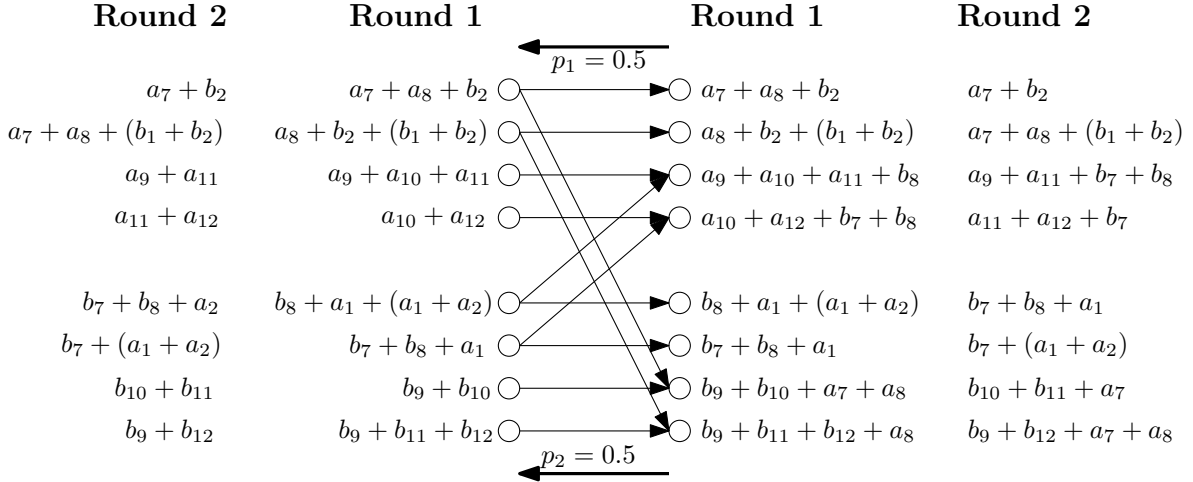


Fig. 8. Second block of transmissions for the example coding scheme over linear deterministic channel. The helping information sent by the interfering transmitters ( $a_1, a_2$  at Rx1,  $b_1, b_2$  at Rx2) are omitted for brevity. Note that these are already known at the receivers from previous block, and hence can be cancelled.

the other hand, for strong interference, rate-limited feedback imposes a hard limit on the amount of gain that can be obtained from feedback, but the gain is still unbounded for intermittent feedback, since the end-to-end mutual information of the alternative path created by feedback gets larger with increasing interference.

#### IV. MOTIVATION OF THE CODING SCHEME

In this section, we illustrate our coding scheme through an example over the linear deterministic channel. This example is intended to demonstrate how and why the proposed scheme works, and motivate the use of quantize-map-forward as a feedback strategy.

We consider the symmetric channel shown in Figures 7 and 8, with  $n_{11} = n_{22} = 4$ ,  $n_{12} = n_{21} = 2$ , and  $p_1 = p_2 = 0.5$ , and focus on the achievable symmetric rate. In this example we will take a block length of  $N = 2$  for illustration purposes. Although for this particular case, the probability of decoding error is large due to short block length, in general the same

coding idea can be applied for a large block length, in which case arbitrarily small error probability can be achieved by taking advantage of the law of large numbers.

We focus on two blocks of transmission. At each block, the users split their messages into common and private parts. The common parts of the messages are decoded by both receivers, whereas the private part is only decoded by the intended receiver, as in Han-Kobayashi scheme for the interference channel without feedback [13]. In the first block, Tx1 sends linear combinations of its two common information symbols,  $a_1, a_2$  on its two common (upper) levels, and linear combinations of its private information symbols,  $a_3, a_4, a_5, a_6$ , over its private (lower) two levels over a block of two time slots. Tx2 performs similar operations for its common symbols  $b_1, b_2$ , and its private symbols  $b_3, b_4, b_5, b_6$ .

Note that at this point, the receivers can decode the symbols sent at their upper two levels by solving the four equations in two unknowns.

After each time slot, the receivers feed back their channel outputs, but the transmitters wait until the end of the block



to collect sufficient information from feedback. We consider a particular feedback channel realization  $(S_1^N, S_2^N) = ((1, 0), (0, 1))$  for illustration purposes. After the first block, each transmitter gets from feedback two linear combinations of the interfering symbols of the previous block, by subtracting their own linear combinations from the channel outputs. In the second block, the transmitters perform further linear encoding of these two linear combinations. These additional linear combinations of the interference symbols are superimposed on top of the linear combinations of the fresh common information symbols  $a_7, a_8$  (and  $b_7, b_8$  for Tx2) of the second block. On the private levels, linear combinations of new symbols  $a_9, a_{10}, a_{11}, a_{12}$  at Tx1 and  $b_9, b_{10}, b_{11}, b_{12}$  at Tx2 are sent, as in the first block.

After the second block of transmission, the receivers collect the four linear equations obtained in the lower two levels of the first block and the four linear equations obtained at the upper two uninterfered levels in the second block. It is easy to check that these eight equations are linearly independent, and hence the receivers can solve for the eight unknowns ( $a_3, a_4, a_5, a_6, a_7, a_8, b_1, b_2$  for Tx1, and  $b_3, b_4, b_5, b_6, b_7, b_8, a_1, a_2$  for Tx2).

Having decoded the private information (and interference) of the first block and the common information of the second block, the receivers next cancel the additional linear combinations of the previously decoded common information received at the lower two levels of the second block due to feedback. This means that Rx1 cancels the  $a_1$  and  $a_2$  symbols in the lower two levels, and Rx2 cancels the  $b_1$  and  $b_2$  symbols.

Since the transmitters can also cancel this information from the received feedback (because it is a function of their own symbols), the state of each terminal reduces to that in the end of the first block. Therefore, in each of the following blocks, the operation in the second block can be repeated, each time letting the receivers decode the private information of the previous block and the common information of the new block.

One caveat is that, the feedback channel realization will not be the same at each block. To address this point, we first note that the only decoding error event is when the channel realization is such that the resulting linear system in any of the receivers is not full rank. For the particular code in the example, it is easy to check that the probability of this event is zero for any feedback channel realization as long as  $S_i^N \neq (0, 0)$  for  $i = 1, 2$ . In general, for any  $\epsilon > 0$ , in order to achieve a symmetric rate  $C_{\text{sym}} - \epsilon$ , Tx $i$  needs to receive feedback for at least  $N(p_i - \epsilon)$  time slots at each block. This condition is ensured by law of large numbers by letting  $N \rightarrow \infty$ , and arbitrarily small error probability can be achieved<sup>6</sup>.

To find the symmetric rate achieved by this scheme, we assume the scheme is run for  $B$  blocks. At the end, each receiver will have resolved  $6B - 4$  information bits in  $2B$  time slots. Letting  $B \rightarrow \infty$  gives a symmetric rate of 3 bits/time slot. Note that without feedback, a symmetric rate of at most 2

bits/time slot can be achieved. At the other extreme, it is also easy to verify from the results in [1] that symmetric capacity under perfect feedback is also 3 bits/time slot, which is in agreement with Figure 6 and Corollary 3.1.

This example also serves to demonstrate why we perform quantize-map-forward instead of decode-and-forward as a feedback strategy. In general, to achieve the symmetric capacity, Tx2 needs to send linear combinations of  $N$  information symbols on its common levels, while Tx1 receives  $2Np_1$  of these linear combinations on the average. Hence, if  $p_1 < 0.5$ , Tx1 will not be able to decode the interference of the previous block. Instead, Tx1 performs a linear mapping of the received feedback information, which turns out to achieve the symmetric capacity.

Finally, we point out that decoding in this scheme is *sequential*, i.e., the receiver decodes the blocks in the same order they are encoded<sup>7</sup>. This is in contrast to earlier feedback coding schemes proposed for interference channel, which perform backward coding. The obvious advantage of using sequential decoding is better delay performance, since the receiver does not need to wait for the end of the entire transmission to start to decode.

## V. ACHIEVABILITY PROOF

In this section, we describe the coding scheme in detail and derive an inner bound  $\mathcal{R}_G^i(p_1, p_2)$  on the rate region.

### A. Overview of the Achievable Strategy

The main idea of the coding scheme is the same as the one presented for the example in Section IV. However, it substantially generalizes the example scheme in order to account for possible channel noise, different interference regimes and an arbitrary target rate point in the achievable region.

The scheme consists of transmission over  $B$  blocks, each of length  $N$ . At the beginning of block  $b$ , upon reception of feedback, transmitters first remove their own contribution from the feedback signal and obtain a function of the interference and noise realization of block  $b - 1$ . This signal is then quantized and mapped to a random codeword, which will be called the helping information. Finally, a new common codeword, which is to be decoded by both receivers, and a private codeword, to be decoded by only the intended receiver, are superimposed to the helping information, and transmitted.

The decoding operation depends on the desired rate point (see Figure 10). To achieve the rate points for which the common component of the message is large, the receiver simply performs a variation of Han-Kobayashi decoding [13], i.e., it decodes the intended information jointly with the common part of the interference. Note that this does not make use of the helping information.

To achieve the remaining rate points, the helping information is used. For weak interference, at block  $b$ , we assume that the receiver has already decoded the intended common information of block  $b - 1$ . After receiving the transmission

<sup>6</sup>Note that this does not prove the existence of a sequence of codes that allows arbitrarily small error probability for an arbitrary block length. The intention in this section is to give an illustration of the coding scheme; the precise achievability proof will be presented in Section V.

<sup>7</sup>An alternate scheme based on backward decoding was presented in [14], for the case of linear deterministic channel.



of block  $b$ , the receivers jointly decode the intended private information and the interference of block  $b - 1$  jointly with the common information of block  $b$ , while using the helping information sent at block  $b$  as side information. For strong interference, the roles of intended common information and the interfering common information get switched.

Next, we present a detailed description of the coding scheme and proof of achievability.

### B. Codebook Generation

Fix  $p(x_{ie})p(x_{ic})p(x_{ip})$  for  $i = 1, 2^8$ , and  $p(u_i|\tilde{v}_j)$  that achieves  $\mathbb{E}[d(U_i, \tilde{V}_j)] \leq D_i$  for  $(i, j) = (1, 2), (2, 1)$ , where  $d: \mathcal{U} \times \mathcal{V} \rightarrow \mathbb{R}$  is the distortion measure, where  $\mathcal{U}$  and  $\mathcal{V}$  are the alphabets of  $U_i$  and  $\tilde{V}_j$ , respectively. Generate  $2^{Nr_i}$  quantization codewords  $U_i^N$  i.i.d.  $\sim p(u_i) = \sum_{\tilde{v}_j} p(u_i|\tilde{v}_j)p(\tilde{v}_j)$ , for  $(i, j) = (1, 2), (2, 1)$ . For  $i = 1, 2$ , generate  $2^{Nr_i}$  codewords  $X_{ie}^N$  i.i.d.  $\sim p(x_{ie})$ . Further generate, for  $i = 1, 2$ ,  $2^{NR_{ic}}$  codewords  $X_{ic}^N$  i.i.d.  $\sim p(x_{ic})$  and  $2^{NR_{ip}}$  codewords  $X_{ip}^N$  i.i.d.  $\sim p(x_{ip})$ . For  $i = 1, 2$ , define symbol-by-symbol mapping functions  $x_i: \mathcal{X}_{if} \times \mathcal{X}_{ip} \rightarrow \mathcal{X}_i$  and  $x_{if}: \mathcal{X}_{ie} \times \mathcal{X}_{ic} \rightarrow \mathcal{X}_{if}$ , where  $\mathcal{X}_{ie}, \mathcal{X}_{ic}, \mathcal{X}_{ip}$ , and  $\mathcal{X}_{if}$  are the alphabets for the symbols  $X_{ie}, X_{ic}, X_{ip}$ , and  $X_{if}$ , respectively.

### C. Encoding

Encoding is performed over blocks (indexed by  $b$ ) of length  $N$ . See Figure 9 for a system diagram. At the beginning of block  $b$ , Tx $i$  receives the punctured feedback signal  $\tilde{Y}_i^N(b-1) = S_i^N(b-1)Y_i^N(b-1)$  containing information about the channel output in block  $b-1$ , where the multiplication is element-wise. Upon reception of  $\tilde{Y}_i^N$ , Tx $i$  first removes its own contribution from the feedback signal to obtain  $\tilde{V}_j^N(b-1) = S_i^N(b-1)V_j^N(b-1)$ . For linear deterministic model, this is done by

$$\tilde{V}_j^N(b-1) = \tilde{Y}_i^N(b-1) - S_i^N(b-1)\mathbf{H}_{ii}X_i^N(b-1),$$

whereas for Gaussian model, it can be obtained by

$$\tilde{V}_j^N(b-1) = \tilde{Y}_i^N(b-1) - S_i^N(b-1)h_{ii}X_i^N(b-1)$$

for  $(i, j) = (1, 2), (2, 1)$ .

The interference signal  $\tilde{V}_j^N(b-1)$  is then quantized by finding an index  $Q_i(b)$  such that

$$(\tilde{V}_j^N(b-1), U_i^N(Q_i(b))) \in \mathcal{T}_\epsilon^{(N)},$$

where  $\mathcal{T}_\epsilon^{(N)}$  denotes the  $\epsilon$ -typical set with respect to the distribution  $p(\tilde{v}_j)p(u_i|\tilde{v}_j)$ , and  $p(\tilde{v}_j)$  is induced by the channel and the input distributions. If such an index  $Q_i(b)$  has been found, the codeword  $X_{ie}^N(Q_i(b))$  that has the same index is chosen to be sent for block  $b$ . If there are multiple such indices, the smallest one is chosen. If no such index is found, the quantization index 1 is chosen.

Next, the message  $W_i(b) \in [2^{NR_i}]$  to be sent at block  $b$  is split into common and private components  $(W_{ic}(b), W_{ip}(b)) \in [2^{NR_{ic}}] \times [2^{NR_{ip}}]$ . Depending on the desired message indices

<sup>8</sup>Although the scheme loses beamforming gain by generating independent codebooks at the two users, this only results in a constant rate penalty.

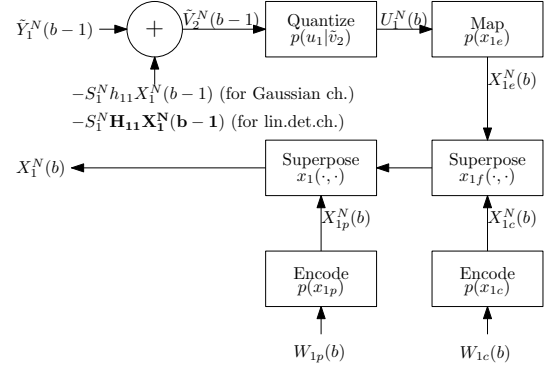


Fig. 9. Encoder diagram at Tx1

$(W_{ic}(b), W_{ip}(b))$ , a common codeword  $X_{ic}^N(W_{ic}(b))$ , and a private codeword  $X_{ip}^N(W_{ip}(b))$  is chosen from the respective codebooks.

Finally, the using the symbol-wise maps  $x_{if}(\cdot, \cdot)$  and  $x_i(\cdot, \cdot)$ , we obtain the codewords

$$\begin{aligned} X_{if}^N(b) &= x_{if}(X_{ie}^N(b), X_{ic}^N(b)) \\ X_i^N(b) &= x_i(X_{if}^N(b), X_{ip}^N(b)) \end{aligned}$$

where the functions are applied to vectors element-wise.  $X_i^N(b)$  is sent at Tx $i$  over  $N$  channel uses.

### D. Decoding

The message indices for common and private messages, and the quantization indices of Tx $i$  at block  $b$  will be denoted by  $m_i(b)$ ,  $n_i(b)$ , and  $q_i(b)$ , respectively. When there are two quantization indices to be decoded from the same user, the second one will be denoted with  $q'_i(b)$ .

In order to describe the decoding process, we need to introduce some notation. Define the following sequence of sets:

$$\begin{aligned} \mathcal{B}_i^{(N)}((q_j, m_j)(b-1)) &:= \left\{ q_i(b) : (\underline{S}^N(b-1), \right. \\ &\quad \left. X_{jf}^N((q_j, m_j)(b-1)), (U_i^N, X_{ie}^N)(q_i(b))) \in \mathcal{T}_\epsilon^{(N)} \right\}. \end{aligned}$$

for  $(i, j) = (1, 2), (2, 1)$ . Loosely,  $\mathcal{B}_i^{(N)}$  is the set of quantization indices of Tx $i$  that are jointly typical with the interference of the previous round. If any of the indices  $(q_j, m_j)$  is known, we will suppress the dependence to that index, e.g., if both are known, we simply denote

$$\begin{aligned} \mathcal{B}_i^{(N)}(b) &:= \left\{ q_i(b) : (\underline{S}^N(b-1), X_{jf}^N(b-1), \right. \\ &\quad \left. (U_i^N, X_{ie}^N)(q_i(b))) \in \mathcal{T}_\epsilon^{(N)} \right\} \end{aligned}$$

where  $X_{jf}^N(b-1)$  refers to the codeword corresponding to the known message indices.

We assume that the set  $\mathcal{B}_i^{(N)}(b)$  has cardinality  $2^{NK_i(b)}$ . Specifically,

$$K_i(b) = \frac{\log \left| \left\{ q_i(b) : (\tilde{V}_j^N(b-1), U_i^N(q_i(b))) \in \mathcal{T}_\epsilon^{(N)} \right\} \right|}{N}$$

Interference Regime	Operating Point	Decoding Operation
Weak Interference	$R_{1c} < I_{w1}$	Jointly decode $W_{1p}(b-1), W_{2c}(b-1), W_{1c}(b)$ , and quantization index $Q_1(b)$
	$R_{1c} \geq I_{w1}$	Jointly decode $W_{1p}(b), W_{1c}(b)$ , and $W_{2c}(b)$ (do not use helping information)
Strong Interference	$R_{2c} < I_{s1}$	Jointly decode $W_{1p}(b-1), W_{1c}(b-1), W_{2c}(b)$ , and quantization index $Q_2(b)$
	$R_{2c} \geq I_{s1}$	Jointly decode $W_{1p}(b), W_{1c}(b)$ , and $W_{2c}(b)$ (do not use helping information)

Fig. 10. A high-level summary of the decoding policy at Rx1 (Details are omitted).

Note that due to random codebook generation,  $K_i(b), i = 1, 2$ , are random variables. The following lemma shows that  $K_i(b)$  is almost surely bounded for sufficiently large  $N$ .

**Lemma 5.1.** *For any  $\epsilon > 0$ , there exists a block length  $N$ , and a quantization scheme such that  $K_i(b) < \kappa_i + \delta(\epsilon)$ , where*

$$\kappa_i := I(\tilde{V}_j; U_i | S_i) - I(X_{jf}; U_i | S_i)$$

for  $(i, j) = (1, 2), (2, 1)$ , and  $\delta(\epsilon)$  is such that  $\delta(\epsilon) \rightarrow 0$  as  $\epsilon \rightarrow 0$ .

*Proof.* See Appendix A.  $\square$

Lemma 5.1 suggests that for each interference codeword, there is a constant number of plausible quantization codewords, for sufficiently large block length (to see that  $\kappa_i$  is a constant independent of channel parameters, refer to Appendix C). This means that the cost of jointly decoding the quantization indices together with the actual messages is a constant reduction in the achievable rate, which will be a useful observation in deriving the constant-gap result.

We also define  $C_i = \kappa_i + 2\kappa_j$ , for  $(i, j) = (1, 2), (2, 1)$ . The reason for this particular definition will become clear in the error analysis. Intuitively,  $C_i$  represents the rate cost associated with performing quantization to forward the feedback information, which introduces distortion. However, as we will show later in the proof, the upper bound given in Lemma 5.1 can be evaluated as a constant independent of channel parameters.

Given an input distribution, Rx1 is said to be in weak interference if  $I(X_2; Y_1 | X_1) \leq I(X_1; Y_1 | X_2)$ , and in strong interference otherwise. These regimes are defined similarly for Rx2.

Decoding operation depends on the interference regime and the desired operating point  $(R_1, R_2)$ . In order to describe the relevant regimes of operating points, we define

$$I_{wi} := I(X_{if}; Y_i | X_{1e}, X_{2e}) - C_i, \quad (16)$$

$$I_{si} := I(X_{jf}; Y_i | X_{1e}, X_{2e}) - C_i, \quad (17)$$

for  $(i, j) = (1, 2), (2, 1)$ . In what follows, for clarity, we will focus only on Rx1. The operations performed at Rx2 are similar.

1) *Weak Interference* ( $I(X_2; Y_1 | X_1) \leq I(X_1; Y_1 | X_2)$ ): If, for the desired operating point,  $R_{1c} > I_{w1}$ , where  $I_{w1}$  is as defined in (16), the helping information is not used, and a slight modification of Han-Kobayashi scheme is employed. Otherwise, the helping information is used to decode the

information of block  $b-1$ . We describe the decoding for the two cases below.

**$R_{1c} \geq I_{w1}$  :** At block  $b$ , we assume that  $X_{1e}^N(b)$  and  $X_1^N(b-1)$  are known. The decoder attempts to find unique indices  $(m_1(b), n_1(b), m_2(b)) \in [2^{NR_{1c}}] \times [2^{NR_{1p}}] \times [2^{NR_{2c}}]$ , and some  $q_2(b) \in [2^{Nr_2}]$  such that

$$\left( \begin{array}{l} \underline{S}^N(b-1), X_{1f}^N(b-1), X_{1e}^N(b), X_{2e}^N(q_2(b)), \\ X_{1f}^N(m_1(b)), X_1^N(m_1(b), n_1(b)), \\ X_{2f}^N(q_2(b), m_2(b)), Y_1^N(b) \end{array} \right) \in \mathcal{T}_\epsilon^{(N)} \quad (18)$$

where the known message indices are suppressed. If the receiver can find a unique collection of such indices, it declares them as the decoded message indices  $(\widehat{W}_{1c}(b), \widehat{W}_{1p}(b), \widehat{W}_{2c}(b))$ ; otherwise it declares an error.

After decoding, given the knowledge of  $X_1^N(b)$ , Rx1 reconstructs  $X_{1e}^N(b+1)$  by imitating the steps taken by Tx1 at the beginning of block  $b+1$ , thereby maintaining the assumption that  $X_{1e}^N(b)$  is known at the beginning of block  $b$ . Further, note that  $X_{2e}^N(b)$  is not uniquely decoded, hence in block  $b+1$ , it will still be jointly (but still, non-uniquely) decoded with the variables of that block. We resort to non-unique decoding of this codeword since unique decoding imposes an additional rate constraint on the helping information, thereby limiting the amount of rate enhancement it can provide.

**$R_{1c} < I_{w1}$  :** At block  $b$ , it is assumed that  $X_{1f}^N(b-1)$  and  $X_1^N(b-2)$  are known at Rx1.

To decode, Rx1 attempts to find unique indices  $(m_1(b), n_1(b-1), m_2(b-1)) \in [2^{NR_{1c}}] \times [2^{NR_{1p}}] \times [2^{NR_{2c}}]$  and some triple  $(q_2(b-1), q_2(b), q_1(b)) \in [2^{Nr_2}] \times [2^{Nr_2}] \times [2^{Nr_1}]$  such that

$$\left( \begin{array}{l} \underline{S}^N(b-1), X_{1f}^N(b-2), X_{1f}^N(b-1), \\ X_1^N(n_1(b-1)), X_{2e}^N(q_2(b-1)), Y_1^N(b-1), \\ (U_1^N, X_{1e}^N(q_1(b)), X_{2e}^N(q_2(b)), \\ X_{2c}^N(m_2(b-1)), X_{1c}^N(m_1(b)), Y_1^N(b) \end{array} \right) \in \mathcal{T}_\epsilon^{(N)} \quad (19)$$

If a unique collection of such indices exists, then these are declared as the decoded message indices  $(\widehat{W}_{1c}(b), \widehat{W}_{1p}(b-1), \widehat{W}_{2c}(b-1))$ . Otherwise, an error is declared.

In (19), the dependence of  $X_1^N(b-1)$  to the indices  $q_1(b-1)$  and  $m_1(b-1)$  is suppressed, since these indices correspond to messages that have already been decoded.

In words, the decoder jointly decodes the private information and the interference of block  $b-1$  jointly with the helping information and common information from block  $b$ .

Note that non-unique decoding is performed for  $X_{1e}^N(b)$ , but we have assumed that  $X_{1f}^N(b-1)$  (and thus,  $X_{1e}^N(b-1)$ ) is uniquely known at the beginning of block  $b$ . In order to maintain this assumption for the next block,  $X_{1e}^N(b)$  is reconstructed at Rx1. To achieve this, given the knowledge of  $X_1^N(b-1)$ , and the quantization codebook, Rx1 imitates the operations performed by Tx1 at the beginning of block  $b$ .

2) *Strong Interference* ( $I(X_2; Y_1 | X_1) > I(X_1; Y_1 | X_2)$ ): As in the weak interference case, decoding depends on the

operating point. For  $R_{2c} < I_{s1}$ , where  $I_{s1}$  is as defined in (17), helping information is used, otherwise, it is not used.

**R<sub>2c</sub> ≥ I<sub>s1</sub>** : The operations performed are identical to those for the case of  $R_{1c} ≥ I_{w1}$  under weak interference.

**R<sub>2c</sub> < I<sub>s1</sub>** : We assume  $X_1^N(b-2)$ ,  $X_{1e}^N(b-1)$ , and  $X_{2c}^N(b-1)$  are known at Rx1 at block  $b$ .

To decode, Rx1 attempts to find unique indices  $(m_1(b-1), n_1(b-1), m_2(b)) \in [2^{NR_{1c}}] \times [2^{NR_{1p}}] \times [2^{NR_{2c}}]$  and some  $(q_2(b-1), q_2(b), q_1(b)) \in [2^{Nr_2}] \times [2^{Nr_2}] \times [2^{Nr_1}]$  such that

$$\begin{pmatrix} S^N(b-1), X_{1f}^N(b-2), X_{1e}^N(b-1), \\ X_{2e}^N(q_2(b-1)), X_{1c}^N(m_1(b-1)), \\ X_{1p}^N(n_1(b-1)), X_{2c}^N(m_2(b)), X_{1e}^N(q_1(b)), \\ (U_2^N, X_{2e}^N(q_2(b)), Y_1^N(b-1), Y_1^N(b)) \end{pmatrix} \in \mathcal{T}_\epsilon^{(N)} \quad (20)$$

If a unique collection of such indices exists, they are declared as the decoded message indices  $(\widehat{W}_{1c}(b-1), \widehat{W}_{1p}(b-1), \widehat{W}_{2c}(b))$ . Otherwise, an error is declared. Using the information of  $X_1^N(b-1)$ , Rx1 can now uniquely reconstruct  $X_{1e}^N(b-1)$  by following the steps taken by Tx1 at the beginning of block  $b$ .

#### E. Error Analysis

Without loss of generality, we only consider the error events occurring at Tx1 and Rx1. All arguments here will be applicable to the other Tx-Rx pair. We define the following decoding error events at Rx1, for block  $b$  and block length  $N$ :

$$\begin{aligned} D_{FB,w}(b, N) &= \left\{ \widehat{W}_{1c}(b) = W_{1c}(b), \right. \\ &\quad \left. \widehat{W}_{1p}(b-1) = W_{1p}(b-1), \widehat{W}_{2c}(b-1) = W_{2c}(b-1) \right\}^c \\ D_{FB,s}(b, N) &= \left\{ \widehat{W}_{1c}(b-1) = W_{1c}(b-1), \right. \\ &\quad \left. \widehat{W}_{1p}(b-1) = W_{1p}(b-1), \widehat{W}_{2c}(b) = W_{2c}(b) \right\}^c \\ D_{NFB}(b, N) &= \left\{ \widehat{W}_1(b) = W_1(b), \widehat{W}_{2c}(b) = W_{2c}(b) \right\}^c \end{aligned}$$

The overall decoding error events at Rx1 is given by

$$\begin{aligned} D_{FB,w}(N) &= \bigcup_{b=1}^B D_{FB,w}(b), \quad D_{FB,s}(N) = \bigcup_{b=1}^B D_{FB,s}(b) \\ D_{NFB}(N) &= \bigcup_{b=1}^B D_{NFB}(b) \end{aligned}$$

We first prove that in order to find the rate achieved after transmission of  $B$  blocks, it is sufficient to focus on the error events at an arbitrary block  $b$ . Without loss of generality, consider the error event  $D_{FB,w}(N)$ . Assume that, after  $B$  blocks of transmission, the effective rate achieved by Tx1 is  $\bar{R}_i$  (Note that at the end of block  $B$ , some of the information pertaining to block  $B$  is still undecoded), which can be lower bounded by  $\bar{R}_i \geq \frac{B-2}{B} R_i$ , by ignoring the partial information decoded in the first block and the last one. We can also upper bound the overall probability of error by

$$\mathbb{P}(D_{FB,w}) \leq \sum_{b=2}^B \mathbb{P}\left(D_{FB,w}(b, N) \mid \{D_{FB,w}^c(b', N)\}_{b'=2}^{b-1}\right)$$

$$\begin{aligned} &\leq B\mathbb{P}\left(D_{FB,w}(b, N) \mid \{D_{FB,w}^c(b', N)\}_{b'=2}^{b-1}\right) \\ &=: B\mathbb{P}(D_{FB,w}(b, N)) \end{aligned}$$

for an arbitrary block  $b$ , where the second line follows by the fact that the encoding and decoding processes are identical in each block, and we made a definition in the last line for brevity<sup>9</sup>. Setting  $B = N = N'$ , we see that for any  $N'$ , an error probability less than  $N'\mathbb{P}(D_{FB,w}(b, N'))$  can be achieved with rate  $\frac{N'-1}{N'} R_i$ . Therefore, in order to show that rate  $R_i$  is achievable, it is sufficient to show that  $N\mathbb{P}(D_{FB,w}(b, N)) \rightarrow 0$  as  $N \rightarrow \infty$ . Using the same arguments, one can show the same result for  $D_{FB,s}(N)$  and  $D_{NFB}(N)$ , and define  $\mathcal{D}_{FB,s}(b, N)$  and  $\mathcal{D}_{NFB}(b, N)$  similarly.

Now we analyze the weak and strong interference regimes separately.

1) *Weak Interference*: The following lemmas characterize the rate constraints for reliable communication with Rx1 for feedback and non-feedback strategies, respectively, under weak interference.

**Lemma 5.2.** *For weak interference at Rx1,  $N\mathbb{P}(D_{FB,w}(b, N)) \rightarrow 0$  as  $N \rightarrow \infty$  if*

$$R_{1c} < I(X_{1f}; Y_1 | \underline{S}, X_{1e}, X_{2e}) - C_1 \quad (21)$$

$$R_{1p} < I(X_1; Y_1 | \underline{S}, X_{1f}, X_{2f}) - C_1 \quad (22)$$

$$R_{2c} < I(X_{2f}; Y_1 | \underline{S}, X_{2e}, X_1) - C_1 \quad (23)$$

$$\begin{aligned} R_{1p} + R_{2c} &< \min \left\{ I(X_1, X_{2f}; Y_1, U_1 | \underline{S}, X_{1f}, X_{2e}) - 2C_1, \right. \\ &\quad \left. I(X_1, X_{2f}; Y_1 | \underline{S}, X_{1c}, X_{2e}) - C_1 \right\} \end{aligned} \quad (24)$$

$$R_1 + R_{2c} < I(X_1, X_{2f}; Y_1 | \underline{S}, X_{1e}, X_{2e}) - C_1 \quad (25)$$

*Proof.* See Appendix B.  $\square$

**Lemma 5.3.** *For weak interference at Rx1,  $N\mathbb{P}(D_{NFB}(b, N)) \rightarrow 0$  as  $N \rightarrow \infty$  if*

$$R_{1c} > I(X_{1f}; Y_1 | \underline{S}, X_{1e}, X_{2e}) - C_1 \quad (26)$$

$$R_{1p} < I(X_1; Y_1 | \underline{S}, X_{1f}, X_{2f}) - \kappa_2 \quad (27)$$

$$R_{2c} < I(X_{2f}; Y_1 | \underline{S}, X_{2e}, X_1) - \kappa_2 \quad (28)$$

$$R_1 < I(X_1; Y_1 | \underline{S}, X_{2f}, X_{1e}) - \kappa_2 \quad (29)$$

$$R_1 + R_{2c} < I(X_1, X_{2f}; Y_1 | \underline{S}, X_{1e}, X_{2e}) - C_1 - \kappa_2 \quad (30)$$

*Proof.* See Appendix B.  $\square$

2) *Strong Interference*: The following lemmas give the rate constraints for the feedback and non-feedback modes under strong interference at Rx1.

**Lemma 5.4.** *For strong interference at Rx1,  $N\mathbb{P}(D_{FB,s}(b, N)) \rightarrow 0$  as  $N \rightarrow \infty$  if*

$$R_{2c} < I(X_{2f}; Y_1 | \underline{S}, X_{1e}, X_{2e}) - C_1 \quad (31)$$

$$R_{1p} < I(X_1; Y_1 | \underline{S}, X_{1f}, X_{2f}) - C_1 \quad (32)$$

$$R_1 < \min \{ I(X_1; Y_1, U_2 | \underline{S}, X_{1e}, X_{2f}), \quad (33)$$

$$I(X_1, X_{2e}; Y_1 | \underline{S}, X_{1e}, X_{2c}) \} - C_1 \quad (34)$$

<sup>9</sup>The event  $\mathcal{D}_{FB,w}$  is defined in the filtered probability space formed by the conditioning.

$$R_1 + R_{2c} < I(X_1, X_{2f}; Y_1 | \underline{S}, X_{1e}, X_{2e}) - C_1 \quad (35)$$

*Proof.* See Appendix B.  $\square$

**Lemma 5.5.** *For strong interference at Rx1,  $N\mathbb{P}(\mathcal{D}_{NFB}(b, N)) \rightarrow 0$  as  $N \rightarrow \infty$  if*

$$R_{2c} > I(X_{2f}; Y_1 | \underline{S}, X_{1e}, X_{2e}) - C_1 \quad (36)$$

$$R_{1p} < I(X_1; Y_1 | \underline{S}, X_{1f}, X_{2f}) - \kappa_2 \quad (37)$$

$$R_{1p} + R_{2c} < I(X_1, X_{2f}; Y_1 | \underline{S}, X_{1f}, X_{2e}) - \kappa_2 \quad (38)$$

$$R_1 + R_{2c} < I(X_1, X_{2f}; Y_1 | \underline{S}, X_{1e}, X_{2e}) - C_1 - \kappa_2 \quad (39)$$

*Proof.* See Appendix B.  $\square$

### F. Rate Region Evaluation

In this subsection, we first explicitly derive the set of achievable  $(R_1, R_2)$  pairs for linear deterministic and Gaussian models, from the results of the previous subsection.

We first find the conditions for decodability at Rx1 under weak interference. Recall that feedback mode is used at Rx1 only if (21) is satisfied; otherwise Han-Kobayashi decoding is performed. If we define  $\underline{R} := (R_{1c}, R_{2c}, R_{1p})$ , and

$$\mathcal{R}_{FB}^w := \{\underline{R} : (22)-(25) \text{ is satisfied}\},$$

$$\mathcal{R}_{NFB}^w := \{\underline{R} : (27)-(30) \text{ is satisfied}\},$$

$$\mathcal{R}_d^w := \{\underline{R} : (21) \text{ is satisfied}\},$$

then the set of rate points  $\mathcal{R}^w$  that ensure decodability at Rx1 under weak interference contains

$$\begin{aligned} \mathcal{R}^w &= (\mathcal{R}_{FB}^w \cap \mathcal{R}_d^w) \cup (\mathcal{R}_{NFB}^w \cap \mathcal{R}_d^{w,c}) \\ &\supseteq (\mathcal{R}_{NFB}^w \cap \mathcal{R}_{FB}^w \cap \mathcal{R}_d^w) \cup (\mathcal{R}_{NFB}^w \cap \mathcal{R}_{FB}^w \cap \mathcal{R}_d^{w,c}) \\ &= \mathcal{R}_{NFB}^w \cap \mathcal{R}_{FB}^w \end{aligned}$$

where  $\mathcal{R}_d^{w,c}$  is the complement of the set  $\mathcal{R}_d^w$ . Therefore, the rate constraints for decodability at Rx1 for the described strategy for weak interference are given by (22)-(25) and (27)-(30), for all joint distributions  $\prod_{i=1}^2 p(x_{ie})p(x_{ic})p(x_{ip})$ , symbol-wise mappings  $x_{if}(x_{ie}, x_{ic})$ ,  $x_i(x_{if}, x_{ip})$ , and  $p(u_i|\tilde{v}_j)$ ,  $(i, j) = (1, 2), (2, 1)$ , consistent with the distortion constraints.

One can perform the same line of arguments as in the case of weak interference to show that the rate constraints for decodability at Rx1 for strong interference are given by (32)-(35) and (37)-(39), for all joint distributions  $\prod_{i=1}^2 p(x_{ie})p(x_{ic})p(x_{ip})$ , symbol-wise mappings  $x_{if}(x_{ie}, x_{ic})$ ,  $x_i(x_{if}, x_{ip})$ , and  $p(u_i|\tilde{v}_j)$ ,  $(i, j) = (1, 2), (2, 1)$ , consistent with the distortion constraints.

Next, we consider linear deterministic and Gaussian models separately, and derive the achievable rate regions explicitly for both cases.

1) *Rate Region for Linear Deterministic Model:* To obtain the achievable rate region, we first evaluate the mutual information terms with specific input distributions. In particular, we choose the distributions and mappings

$$X_{ie} \sim \text{Unif}[\mathbb{F}_2^{n_{ji}}] \quad (40)$$

$$X_{ic} \sim \text{Unif}[\mathbb{F}_2^{n_{ji}}] \quad (41)$$

$$X_{ip} \sim \text{Unif}[\mathbb{F}_2^{(n_{ii}-n_{ji})^+}] \quad (42)$$

$$U_i = \tilde{V}_j \quad (43)$$

$$x_{if} : \mathbb{F}_2^{n_{ji}} \times \mathbb{F}_2^{n_{ji}} \rightarrow \mathbb{F}_2^{n_{ji}},$$

$$x_i : \mathbb{F}_2^{n_{ji}} \times \mathbb{F}_2^{(n_{ii}-n_{ji})^+} \rightarrow \mathbb{F}_2^{\max(n_{ii}, n_{ji})},$$

$$x_i = [X_{if} \ X_{ip}]^T, \quad x_{if}(a, b) = a + b \quad (44)$$

for  $(i, j) = (1, 2), (2, 1)$ , where  $\text{Unif}[\mathcal{A}]$  denotes uniform distribution over the set  $\mathcal{A}$ . Evaluating the mutual information terms of the previous subsection with this set of distributions, and applying Fourier-Motzkin elimination (see Appendix C for details), we obtain the rate region given in (1)-(6).

2) *Rate Region for Gaussian Model:* Now we evaluate the rate constraints obtained in the previous section, and obtain the final achievable rate region. Assuming available power  $P_i$  at Tx $i$ , we assign the following input distributions, for  $(i, j) = (1, 2), (2, 1)$ :

$$X_{ie} \sim \mathcal{CN}\left(0, \frac{1}{2}P_i\right) \quad (45)$$

$$X_{ic} \sim \mathcal{CN}\left(0, \frac{1}{2}(1 - P_{ip})P_i\right) \quad (46)$$

$$X_{ip} \sim \mathcal{CN}\left(0, \frac{1}{2} \min\left(\frac{1}{|h_{ji}|^2 P_i}, 1\right) P_i\right) \quad (47)$$

$$U_i | \tilde{V}_j \sim \mathcal{CN}(\tilde{V}_j, D_i) \quad (48)$$

$$x_{if} : \mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C}, \quad x_i : \mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C},$$

$$x_{if}(a, b) = a + b, \quad x_i(a, b) = a + b \quad (49)$$

where  $D_i > 0$  are the distortion parameters. Using these input distributions, and applying Fourier-Motzkin elimination (See Appendix C for details), we can show that the rate region (79)-(81), given in Appendix C, is achievable.

## VI. CONVERSE PROOF

We now prove an outer bound region that exactly matches the region given in (1)-(6), and is within a constant gap of the region in (11)-(14).

The main idea between the novel bounds on  $R_1$  and  $R_2$  is based on a genie argument, where the receivers are provided with side-information about the messages. The bounds on  $R_1 + R_2$ ,  $2R_1 + R_2$  and  $R_1 + 2R_2$  are proven through a channel enhancement technique, resembling the one used for the multiple-access channel in [16].

### A. Bounds on $R_1$ and $R_2$

Since any outer bound for perfect feedback is also an outer bound for intermittent feedback, we have the perfect feedback bound

$$R_i \leq \max(n_{ii}, n_{ij}) \quad (50)$$

for linear deterministic model, and the bound

$$R_i \leq \sup_{0 \leq \rho \leq 1} \log\left(1 + \text{SNR}_i + \text{INR}_i + 2\rho\sqrt{\text{SNR}_i \cdot \text{INR}_i}\right) \quad (51)$$

for Gaussian model, for  $(i, j) = (1, 2), (2, 1)$ , which are both proved in [1]. Next, we prove a novel bound for both models.

Without loss of generality, we focus on the bound on  $R_1$ . In order to prove the novel bound on  $R_1$ , the main idea is to provide  $(W_2, \tilde{V}_1^N)$  as side-information to Rx1. The intuition behind this particular choice is revealed when we consider the interference regime and operating point in which this bound is active. First, due to the structure of the capacity region, this bound is relevant only when the message (*i.e.*, the rate) of the interfering user is small enough. Hence, for that regime,  $W_2$  does not carry too much information, and thus providing this to Rx1 still results in a tight outer bound. Second, note that this bound is only active in the strong interference regime, where feedback from Rx2 to Tx2 creates an alternative path for the transmission of  $W_1$ . Therefore, by forwarding this information, Tx2 indeed provides the information contained in  $\tilde{V}_1^N$  to Rx1.

Based on this idea, we prove the bound

$$R_i \leq n_{ii} + p_j (n_{ji} - n_{ii})^+ \quad (52)$$

for linear deterministic model in Appendix D, and the bound

$$R_i \leq \log(1 + \text{SNR}_i) + p_j \log\left(1 + \frac{\text{INR}_j}{1 + \text{SNR}_i}\right) \quad (53)$$

for the Gaussian model in Appendix E, for  $(i, j) = (1, 2), (2, 1)$ .

#### B. Bounds on $R_1 + R_2$ , $2R_1 + R_2$ and $R_1 + 2R_2$

We have the perfect feedback outer bounds

$$R_i + R_j \leq \max(n_{ii}, n_{ij}) + (n_{jj} - n_{ji})^+ \quad (54)$$

for linear deterministic model, and

$$R_i + R_j < \sup_{0 \leq \rho \leq 1} \log\left(1 + \frac{(1 - \rho^2)\text{SNR}_i}{1 + (1 - \rho^2)\text{INR}_j}\right) + \log\left(1 + \text{SNR}_j + \text{INR}_j + 2\rho\sqrt{\text{SNR}_j \cdot \text{INR}_j}\right) \quad (55)$$

for Gaussian model, for  $(i, j) = (1, 2), (2, 1)$ .

Next, we prove novel outer bounds on the capacity region. The novelty in these bounds is in the fact that it combines the existing genie-aided bounding techniques for interference channel with the channel enhancement technique of [16]. In order to prove these bounds, we first define a notion of enhanced channel. Considering our achievable scheme, feedback can be interpreted as a mechanism for the receivers to separate the interference and the intended signal, to the extent allowed by the erasure probability in the feedback channel. In the weak interference regime, this allows the receiver to cancel the interference. In the strong interference regime, through the alternate path created by the interfering user, it allows the reception of additional information about the intended message. Therefore we consider an enhanced channel where the receivers observe the interference and the intended signal individually whenever the feedback is available, and their sum otherwise. In addition to this enhancement, we provide Rx*i* with the side-information of  $V_i^N$  as well, as was done in [17]. To make this more precise, we consider the two models separately.

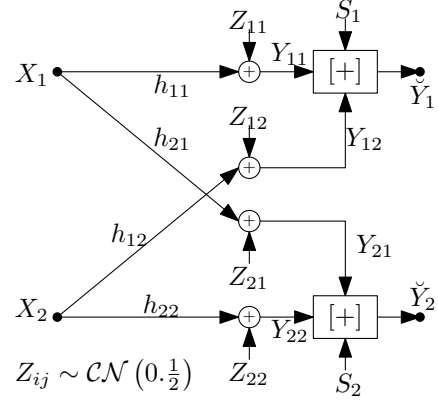


Fig. 11. The enhanced channel for Gaussian model. The block  $[+]$  is a conditional adder, which outputs the sum of the other two inputs if  $S_i = 0$ , and outputs the two inputs separately otherwise.

1) *Linear Deterministic Model:* We define the enhanced linear deterministic channel with intermittent feedback by the following equations

$$\check{Y}_i = \begin{cases} Y_i, & \text{if } S_i = 0 \\ (\mathbf{H}_{ii}X_i, V_j), & \text{if } S_i = 1 \end{cases}$$

for  $(i, j) = (1, 2), (2, 1)$ , where  $\check{Y}_i$  is the channel output of the enhanced channel at Rx*i*,  $Y_i$  is the channel output of the original channel, and  $X_i$  and  $V_j$  are as defined for the original channel. The output of the feedback channel is given by  $\check{Y}_i = S_i Y_i$ , *i.e.*, the same as the original channel. Note that any scheme that achieves arbitrarily small error probability in the original channel can also achieve arbitrarily small error probability for the enhanced channel, using the fact that  $Y_i = \mathbf{H}_{ii}X_i + V_j$ . This means that the capacity region of the original channel is a subset of that of the enhanced channel, and we can derive an outer bound for the enhanced channel instead.

It is easy to see that this enhancement is equivalent to providing the Rx*i* with  $\tilde{V}_2^N$ , since for time slots where  $S_i = 1$ , Rx*i* can use this information to individually obtain the interference and the intended symbol.

Using the channel enhancement technique, we arrive at the following outer bounds on the capacity region of the linear deterministic interference channel with intermittent feedback, which are explicitly proved in Appendix D.

$$R_1 + R_2 \leq \max\{n_{12}, (n_{11} - n_{21})^+\} + \max\{n_{21}, (n_{22} - n_{12})^+\} + p_1 \min\{n_{12}, (n_{11} - n_{21})^+\} + p_2 \min\{n_{21}, (n_{22} - n_{12})^+\} \quad (56)$$

$$2R_1 + R_2 \leq \max(n_{11}, n_{12}) + \max\{n_{21}, (n_{22} - n_{12})^+\} + (n_{11} - n_{21})^+ + p_2 \min\{n_{21}, (n_{22} - n_{12})^+\} \quad (57)$$

$$R_1 + 2R_2 \leq \max(n_{22}, n_{21}) + \max\{n_{12}, (n_{11} - n_{21})^+\} + (n_{22} - n_{12})^+ + p_1 \min\{n_{12}, (n_{11} - n_{21})^+\} \quad (58)$$

2) *Gaussian Model:* Next, we extend the enhanced channel idea to the Gaussian model. In this case, while splitting

the interference and the intended signal, we also split the noise evenly between these two variables (see Figure 11). Specifically, we consider the channel defined by the equations

$$\check{Y}_i = \begin{cases} \bar{Y}_i, & \text{if } S_i = 0 \\ (Y_{ii}, Y_{ij}), & \text{if } S_i = 1 \end{cases}$$

for  $(i, j) = (1, 2), (2, 1)$ , where  $\check{Y}_i$  is the output of the enhanced channel, and

$$\begin{aligned} Y_{ii} &= h_{ii}X_i + Z_{ii} \\ Y_{ij} &= h_{ij}X_j + Z_{ij} \\ \bar{Y}_i &= Y_{ii} + Y_{ij} = h_{ii}X_i + h_{ij}X_j + \bar{Z}_i \end{aligned}$$

with  $Z_{ij}, Z_{ii}$  are independent and distributed with  $\mathcal{CN}(0, \frac{1}{2})$ , and we define  $\bar{Z}_i = Z_{ii} + Z_{ij}$ . The output of the feedback channel at Tx*i* is given by  $S_i \bar{Y}_i = S_i \cdot (Y_{ii} + Y_{ij})$ , i.e., the same as the original channel. It is worth noting that unlike the linear deterministic case, this enhancement is not equivalent to providing Rx*i* with  $\check{V}_j^N$ , since giving this side-information allows the receiver to completely cancel the noise for some time slots, resulting in an infinitely loose bound.

Let  $\mathcal{C}_e(p_1, p_2)$  denote the capacity region of the enhanced channel.

The next lemma shows that the capacity region of the enhanced channel indeed dominates the original one.

**Lemma 6.1.** *For all  $0 \leq p_1, p_2 \leq 1$ ,*

$$\mathcal{C}_G(p_1, p_2) \subseteq \mathcal{C}_e(p_1, p_2)$$

*Proof.* The proof has two steps. First, we consider an intermediate channel, with capacity region  $\mathcal{C}_i(p_1, p_2)$ , and the channel output at Rx*i* is given by

$$Y_i = h_{ii}X_i + h_{ij}X_j + \bar{Z}_i$$

for  $(i, j) = (1, 2), (2, 1)$ , where  $\bar{Z}_i = Z_{ii} + Z_{ij}$  is the sum of two independent  $\mathcal{CN}(0, \frac{1}{2})$  random variables as in the enhanced channel. Since  $\bar{Z}_i$  and  $Z_i$  (the noise in the original channel) have the same probability distribution and are both i.i.d. processes across time and across users, the joint distribution of the channel  $p(y_1, y_2 | x_1, x_2)$  is identical for both channels, and hence they have the same feedback capacity region, i.e.,  $\mathcal{C}_i(p_1, p_2) = \mathcal{C}_G(p_1, p_2)$ .

Next, comparing the intermediate channel and the enhanced channel, we note that any rate pair  $(R_1, R_2)$  achievable in the intermediate channel is also achievable for the enhanced channel using the same pair of codes, using the fact that  $\bar{Y}_i = Y_{ii} + Y_{ij}$ . Therefore,  $\mathcal{C}_i(p_1, p_2) \subseteq \mathcal{C}_e(p_1, p_2)$ , which completes the proof.  $\square$

**Remark 6.1.** *We note that a similar channel enhancement technique has been applied before by Khisti and Lapidoth [16], for Gaussian multiple-access channel with intermittent feedback. In that work, the variances of the random variables  $Z_{ii}$  and  $Z_{ij}$  are not fixed, but are arbitrary, subject to the constraint that they sum to one. Although one can optimize over the noise variances in order to obtain the tightest bound, this only results in a small and constant improvement. Hence, for simplicity, we stick to the fixed variance of  $\frac{1}{2}$  for the noise variables of the enhanced channel.*

Using Lemma 6.1, we can instead prove outer bounds for the enhanced channel. In Appendix E, we prove the following bounds.

$$\begin{aligned} R_1 + R_2 &\leq \log \left( 1 + \text{INR}_1 + \frac{\text{SNR}_1 + 2\sqrt{\text{SNR}_1 \cdot \text{INR}_1}}{1 + \text{INR}_2} \right) \\ &+ \log \left( 1 + \text{INR}_2 + \frac{\text{SNR}_2 + 2\sqrt{\text{SNR}_2 \cdot \text{INR}_2}}{1 + \text{INR}_1} \right) \\ &+ p_1 \log \left( \frac{(1 + 2\text{INR}_1) \left( 1 + \frac{\text{SNR}_1}{\text{INR}_2 + \frac{1}{2}} \right)}{1 + \text{INR}_1 + \frac{\text{SNR}_1 + 2\sqrt{\text{SNR}_1 \cdot \text{INR}_1}}{1 + \text{INR}_2}} \right) \\ &+ p_2 \log \left( \frac{(1 + 2\text{INR}_2) \left( 1 + \frac{\text{SNR}_2}{\text{INR}_1 + \frac{1}{2}} \right)}{1 + \text{INR}_2 + \frac{\text{SNR}_2 + 2\sqrt{\text{SNR}_2 \cdot \text{INR}_2}}{1 + \text{INR}_1}} \right) \end{aligned} \quad (59)$$

$$\begin{aligned} 2R_1 + R_2 &\leq \log \left( 1 + \text{SNR}_1 + \text{INR}_1 + 2\sqrt{\text{SNR}_1 \cdot \text{INR}_1} \right) \\ &+ \log \left( 1 + \frac{\text{SNR}_1}{\frac{1}{2} + \text{INR}_2} \right) \\ &+ \log \left( 1 + \text{INR}_2 + \frac{\text{SNR}_2 + 2\sqrt{\text{SNR}_2 \cdot \text{INR}_2}}{1 + \text{INR}_1} \right) \\ &+ p_2 \log \left( \frac{(1 + 2\text{INR}_2) \left( 1 + \frac{\text{SNR}_2}{\text{INR}_1 + \frac{1}{2}} \right)}{1 + \text{INR}_2 + \frac{\text{SNR}_2 + 2\sqrt{\text{SNR}_2 \cdot \text{INR}_2}}{1 + \text{INR}_1}} \right) \end{aligned} \quad (60)$$

$$\begin{aligned} R_1 + 2R_2 &\leq \log \left( 1 + \text{SNR}_2 + \text{INR}_2 + 2\sqrt{\text{SNR}_2 \cdot \text{INR}_2} \right) \\ &+ \log \left( 1 + \frac{\text{SNR}_2}{\frac{1}{2} + \text{INR}_1} \right) \\ &+ \log \left( 1 + \text{INR}_1 + \frac{\text{SNR}_1 + 2\sqrt{\text{SNR}_1 \cdot \text{INR}_1}}{1 + \text{INR}_2} \right) \\ &+ p_1 \log \left( \frac{(1 + 2\text{INR}_1) \left( 1 + \frac{\text{SNR}_1}{\text{INR}_2 + \frac{1}{2}} \right)}{1 + \text{INR}_1 + \frac{\text{SNR}_1 + 2\sqrt{\text{SNR}_1 \cdot \text{INR}_1}}{1 + \text{INR}_2}} \right) \end{aligned} \quad (61)$$

## VII. DISCUSSION AND EXTENSIONS

We considered the interference channel with intermittent feedback, and derived an approximate characterization of the capacity region under Gaussian model, as well as an exact characterization for the linear deterministic case. The result shows that even intermittent feedback provides multiplicative gain in capacity in interference channels. The achievability result was based on quantize-map-forward relaying at the transmitters, and the outer bound result was based on a channel enhancement technique.

In this paper, we considered short messaging, i.e., a new message is sent at every block of transmission. An alternate approach one could try is long messaging, where the transmitters send codewords describing the same message at every block, and the receivers jointly decode all blocks to recover the message. The clear advantage of short messaging approach is better delay performance, since each message is decoded immediately after the transmission of the corresponding block, instead of waiting for the end of the entire transmission. However, combined with forward decoding, the rate region achievable by this strategy cannot approximate the entire capacity region by itself, as can be seen from the results of Section V; we need to take the union with Han-Kobayashi

rate region to approximate the entire capacity region. This is because while decoding block  $b$ , part of the message of block  $b+1$  is jointly decoded by treating the interference of block  $b+1$  as noise, which limits the rate in certain operating points. Hence, long-messaging approach would remove the need for taking union with Han-Kobayashi region and simplify the proof, since all blocks are jointly decoded. Such an approach has been taken in [18] to derive an inner bound on the capacity region of interference channels with generalized feedback, which overlaps with the capacity region (1)–(6) for the special case of linear deterministic IC with intermittent feedback.

The extension to parallel channels is carried out for the special case of identical subchannels in this work. An important generalization can be the case where the channel gains of the subchannels are not necessarily the same. The main obstacle in generalizing our achievable scheme to this case is that it distinguishes the cases of weak and strong interference, although such a separation is not possible for vector channels. Again, long-messaging can be a strategy to circumvent this issue, since it removes the need for making such a distinction between weak and strong interference regimes [18], albeit at the cost of a much larger delay.

Another important extension could be to the additive white Gaussian noise (AWGN) feedback model of [5]. Since this model assumes passive feedback as well, our quantize-map-forward based scheme can be directly applied to this channel model. The results of this paper indicate that quantize-map-forward, as a feedback strategy, might be a promising candidate as an approximately-capacity-achieving scheme for AWGN feedback model. However, this investigation is not the focus of this paper, and is left as future work.

## REFERENCES

- [1] C. Suh and D. N. C. Tse, "Feedback capacity of the Gaussian interference channel to within 2 bits," *IEEE Transactions on Information Theory*, vol. 57, pp. 2667–2685, May 2011.
- [2] C. E. Shannon, "The zero error capacity of a noisy channel," *Information Theory, IRE Transactions on*, vol. 2, no. 3, pp. 8–19, 1956.
- [3] L. H. Ozarow, "The capacity of the white gaussian multiple access channel with feedback," *Information Theory, IEEE Transactions on*, vol. 30, no. 4, pp. 623–629, 1984.
- [4] A. Vahid, C. Suh, and A. S. Avestimehr, "Interference channels with rate-limited feedback," *IEEE Transactions on Information Theory*, vol. 58, pp. 2788–2812, May 2012.
- [5] S.-Q. Le, R. Tandon, M. Motani, and H. V. Poor, "The capacity region of the symmetric linear deterministic interference channel with partial feedback," *Proceedings of Allerton Conference on Communication, Control, and Computing*, October 2012.
- [6] A. Sahai, V. Aggarwal, M. Yuksel, and A. Sabharwal, "On channel output feedback in deterministic interference channels," in *Information Theory Workshop, 2009. ITW 2009. IEEE*, pp. 298–302, IEEE, 2009.
- [7] C. Suh, I.-H. Wang, and D. N. C. Tse, "Two-way interference channels," *Proceedings of IEEE International Symposium on Information Theory*, pp. 2811–2815, July 2012.
- [8] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: A deterministic approach," *IEEE Transactions on Information Theory*, vol. 57, pp. 1872–1905, April 2011.
- [9] S. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *Information Theory, IEEE Transactions on*, vol. 57, no. 5, pp. 3132–3152, 2011.
- [10] A. Ozgur and S. Diggavi, "Approximately achieving gaussian relay network capacity with lattice codes," *arXiv preprint arXiv:1005.1284*, 2010.
- [11] A. Ozgur and S. Diggavi, "Approximately achieving gaussian relay network capacity with lattice-based qmf codes," *IEEE Transactions on Information Theory*, vol. 59, pp. 8275–8294, Dec 2013.
- [12] G. Kramer and J. Hou, "On message lengths for noisy network coding," in *Information Theory Workshop (ITW), 2011 IEEE*, pp. 430–431, IEEE, 2011.
- [13] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Transactions on Information Theory*, vol. 27, pp. 49–60, January 1981.
- [14] C. Karakus, I.-H. Wang, and S. Diggavi, "Interference channel with intermittent feedback," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pp. 26–30, IEEE, 2013.
- [15] C. Karakus, I.-H. Wang, and S. Diggavi, "An achievable rate region for gaussian interference channel with intermittent feedback," in *Communication, Control, and Computing (Allerton), 2013 51st Annual Allerton Conference on*, pp. 203–210, Oct 2013.
- [16] A. Khisti and A. Lapidoth, "Multiple access channels with intermittent feedback and side information," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pp. 2631–2635, IEEE, 2013.
- [17] R. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Transactions on Information Theory*, vol. 54, pp. 5534–5562, December 2008.
- [18] A. Zaidi, "Achievable regions for interference channels with generalized and intermittent feedback," in *Information Theory Proceedings (ISIT), 2014 IEEE International Symposium on*, pp. 1026–1030, IEEE, 2014.
- [19] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.
- [20] R. Durrett, *Probability: theory and examples*, vol. 3. Cambridge university press, 2010.

## APPENDIX A PROOF OF LEMMA 5.1

Choose  $\epsilon > 0$ . We suppress the dependence of variables on block index  $b$  for simplicity. Consider, for  $(i, j) = (1, 2), (2, 1)$ ,

$$\begin{aligned}\mathbb{E}[2^{NK_i}] &= \mathbb{E}\left[\sum_{q_i=1}^{2^{Nr_i}} \mathbb{1}_{\{q_i: (X_{jf}^N, U_i^N(q_i)) \in \mathcal{T}_\epsilon^{(N)}\}}\right] \\ &= \sum_{q_i=1}^{2^{Nr_i}} \mathbb{P}\left((X_{jf}^N, U_i^N(q_i)) \in \mathcal{T}_\epsilon^{(N)}\right) \\ &= 2^{Nr_i} \mathbb{P}\left((X_{jf}^N, U_i^N(1)) \in \mathcal{T}_\epsilon^{(N)}\right)\end{aligned}$$

Since  $U_i^N(1)$  is generated independently from  $X_{jf}^N$ , by packing lemma [19], there exists  $\delta(\epsilon)$  with  $\delta(\epsilon) \rightarrow 0$  such that

$$\mathbb{E}[2^{NK_i}] \leq 2^{Nr_i} 2^{-N[I(X_{jf}; U_i) - \delta(\epsilon)/3]}$$

for all  $N$ .

Next consider the variance of  $2^{NK_i}$ .

$$\begin{aligned}\text{var}(2^{NK_i}) &= \text{var}\left(\sum_{q_i=1}^{2^{Nr_i}} \mathbb{1}_{\{q_i: (X_{jf}^N, U_i^N(q_i)) \in \mathcal{T}_\epsilon^{(N)}\}}\right) \\ &\stackrel{(a)}{=} \sum_{q_i=1}^{2^{Nr_i}} \text{var}\left(\mathbb{1}_{\{q_i: (X_{jf}^N, U_i^N(q_i)) \in \mathcal{T}_\epsilon^{(N)}\}}\right) \\ &= \sum_{q_i=1}^{2^{Nr_i}} \mathbb{E}\left[\mathbb{1}_{\{q_i: (X_{jf}^N, U_i^N(q_i)) \in \mathcal{T}_\epsilon^{(N)}\}}\right] \\ &\quad \cdot \left(1 - \mathbb{E}\left[\mathbb{1}_{\{q_i: (X_{jf}^N, U_i^N(q_i)) \in \mathcal{T}_\epsilon^{(N)}\}}\right]\right) \\ &= \sum_{q_i=1}^{2^{Nr_i}} \mathbb{P}\left((X_{jf}^N, U_i^N(q_i)) \in \mathcal{T}_\epsilon^{(N)}\right) \\ &\quad \cdot \left(1 - \mathbb{P}\left((X_{jf}^N, U_i^N(q_i)) \in \mathcal{T}_\epsilon^{(N)}\right)\right)\end{aligned}$$



$$\stackrel{(b)}{=} 2^{Nr_i} p_N (1 - p_N) \leq 2^{Nr_i} p_N$$

where (a) is due to independence of the indicator variables, and we have defined  $p_N := \mathbb{P}\left(\left(X_{jf}^N, U_i^N(1)\right) \in \mathcal{T}_\epsilon^{(N)}\right)$  in (b). Hence, there exists  $N_1$  such that for all  $N > N_1$ ,

$$\text{var}(2^{NK_i}) \leq 2^{Nr_i} 2^{-N[I(X_{jf}; U_i) - \delta(\epsilon)]}$$

for some  $\delta(\epsilon)$  with  $\delta(\epsilon) \rightarrow 0$  as  $\epsilon \rightarrow \infty$ .

Define  $\eta := 2^{Nr_i} 2^{-N[I(X_{jf}; U_i) - \delta(\epsilon)/3]} (2^{N\delta(\epsilon)/3} - 1)$ , and the sequence of events

$$\mathcal{E}_n := \left\{ \left| 2^{(n+N_1)K_i} - \mathbb{E} \left[ 2^{(n+N_1)K_i} \right] \right| > \eta \right\}$$

indexed by  $n \geq 1$ .

Borel-Cantelli lemma [20] states that if  $\sum_{n=1}^{\infty} \mathbb{P}(\mathcal{E}_n) < \infty$ , then  $\mathbb{P}(\mathcal{E}_n \text{ infinitely often}) = 0$ . Then consider

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}(\mathcal{E}_n) &= \sum_{n=1}^{\infty} \mathbb{P}\left(\left| 2^{(n+N_1)K_i} - \mathbb{E} \left[ 2^{(n+N_1)K_i} \right] \right| > \eta\right) \\ &\stackrel{(a)}{\leq} \sum_{n=1}^{\infty} \frac{\text{var}(2^{(n+N_1)K_i})}{\eta^2} \\ &\leq \sum_{n=1}^{\infty} \frac{1}{2^{n[r_i - I(X_{jf}; U_i) + \delta(\epsilon)/3]} (2^{n\delta(\epsilon)/3} - 1)^2} \\ &\stackrel{(b)}{<} \infty \end{aligned}$$

where (a) follows by Chebyshev's inequality, and (b) is because exponentially decaying series converge, and  $r_i > I(\tilde{V}_j; U_i) \geq I(X_{jf}; U_i)$  where the first inequality is by covering lemma [19], and the second is by data processing inequality (recall that  $X_{jf} - \tilde{V}_j - U_i$  is a Markov chain). Therefore, with probability one, there exists a finite integer  $N_2 \geq N_1$  such that for all  $N \geq N_2$ ,

$$2^{NK_i} < 2^{Nr_i} 2^{-N[I(X_{jf}; U_i) - 2\delta(\epsilon)/3]}$$

Choosing  $r_i = I(\tilde{V}_j; U_i) + \delta(\epsilon)/3$ , taking the logarithm of both sides, and dividing by  $N$ , we get the desired result.

## APPENDIX B

### PROOFS OF LEMMAS 5.2, 5.3, 5.4 AND 5.5

#### A. Notation

We will often suppress the dependence on block index  $b$  and block length  $N$  for brevity.

For any given set of message indices  $(m_1, n_1, m_2)$ , define the following events, with a little abuse of notation

$$\begin{aligned} T(m_1, n_1, m_2) &:= \{ \exists (q_1, q_2, q'_2) \text{ s.t. (19) holds for the} \\ &\quad \text{indices } (m_1, n_1, m_2, q_1, q_2, q'_2) \}, \\ T(m_1, n_1, m_2, q_1, q_2, q'_2) &:= \{ \text{(19) holds for the indices} \\ &\quad (m_1, n_1, m_2, q_1, q_2, q'_2) \}. \end{aligned}$$

We also define the following quantization error event at Tx<sub>i</sub>

$$\begin{aligned} E_i &= \left\{ \tilde{V}_j^N \in \mathcal{T}_\epsilon^{(N)}, \left( \tilde{V}_j^N, U_i^N(q_i) \right) \notin \mathcal{T}_\epsilon^{(N)} \forall q_i \right\} \\ &\cup \left\{ \tilde{V}_j^N \notin \mathcal{T}_\epsilon^{(N)} \right\} \end{aligned}$$

for  $(i, j) = (1, 2), (2, 1)$ , and  $E := E_1 \cup E_2$ .

Without loss of generality, we assume that the correct message and quantization indices correspond to the index 1, i.e.  $(W_{1c}, W_{1p}, W_{2c}, Q_1, Q_2) = (1, 1, 1, 1, 1)$  for all blocks. We introduce the notation

$$\bar{\mathcal{B}}_i(b) := \mathcal{B}_i(b) \setminus \{1\}.$$

An arbitrary element of the set  $\bar{\mathcal{B}}_i(b)$  will be denoted with  $\bar{q}_i$ , or  $\bar{q}'_i$ . In this analysis, we focus on an arbitrary block  $b$ , but we will also need to refer to variables from block  $b-1$ . The variables associated with block  $b-1$  will be represented with a caron notation when in single letter form. For example, while  $\tilde{X}_{2e}$  is the single letter form for  $X_{2e}^N(b-1)$ ,  $X_{2e}$  is the single letter form for  $X_{2e}^N(b)$ . The feedback state pair  $\underline{S} = (S_1, S_2)$  is assumed to be conditioned upon in all the mutual information terms (since the receivers have access to this information causally), but will be omitted for brevity.

#### B. Claims

In this subsection, we will prove two simple claims that will be useful in bounding the probability of decoding error.

**Claim B.1.** *Let  $A_k$ ,  $k = 1, 2, \dots$  be a sequence of i.i.d. events. Let  $\mathcal{S} \subset \mathbb{N}$  be a random subset of natural numbers (not necessarily independent from the events  $A_k$ ) such that  $|\mathcal{S}| \leq M$  a.s. for some real number  $M$ , and  $\mathbb{P}(A_k|\mathcal{S}) = \mathbb{P}(A_m|\mathcal{S})$  a.s. for all  $(k, m)$  pairs. Then*

$$\mathbb{P}\left(\bigcup_{k \in \mathcal{S}} A_k\right) \leq M \mathbb{P}(A_j)$$

for an arbitrary  $j$ .

*Proof.*

$$\begin{aligned} \mathbb{P}\left(\bigcup_{k \in \mathcal{S}} A_k\right) &= \mathbb{E} \left[ \mathbb{E} \left[ \mathbb{1}_{\bigcup_{k \in \mathcal{S}} A_k} | \mathcal{S} \right] \right] \leq \mathbb{E} \left[ \mathbb{E} \left[ \sum_{k \in \mathcal{S}} \mathbb{1}_{A_k} | \mathcal{S} \right] \right] \\ &\stackrel{(a)}{=} \mathbb{E} \left[ \mathbb{E} \left[ |\mathcal{S}| \mathbb{1}_{A_1} | \mathcal{S} \right] \right] = \mathbb{E} \left[ |\mathcal{S}| \mathbb{E} \left[ \mathbb{1}_{A_1} | \mathcal{S} \right] \right] \\ &\leq \mathbb{E} \left[ M \mathbb{E} \left[ \mathbb{1}_{A_1} | \mathcal{S} \right] \right] = M \mathbb{E} \left[ \mathbb{E} \left[ \mathbb{1}_{A_1} | \mathcal{S} \right] \right] = M \mathbb{P}(A_1) \end{aligned}$$

where (a) follows by the fact that  $\mathbb{P}(A_k|\mathcal{S})$  is the same for all  $k$ .  $\square$

**Claim B.2.** *Let  $(X^N, Y^N, Z^N)$  be distributed i.i.d. according to  $p(x, y, z)$ , and  $(\tilde{X}^N, \tilde{Y}^N, \tilde{Z}^N)$  be distributed i.i.d. according to  $p(x)p(y)p(z)$ . Then there exists  $\delta(\epsilon)$  with  $\lim_{\epsilon \rightarrow 0} \delta(\epsilon) = 0$  such that*

$$\mathbb{P}\left(\left(\tilde{X}^N, \tilde{Y}^N, \tilde{Z}^N\right) \in \mathcal{T}_\epsilon^{(N)}\right) \leq 2^{-N[I(X; Y) + I(Z; X, Y) - \delta(\epsilon)]}$$

*Proof.*

$$\begin{aligned} \mathbb{P}\left(\left(\tilde{X}^N, \tilde{Y}^N, \tilde{Z}^N\right) \in \mathcal{T}_\epsilon^{(N)}\right) &\leq 2^{-N[D(P_{X,Y,Z} || P_X P_Y P_Z) - \delta(\epsilon)]} \\ &= 2^{-N[I(X; Y) + I(Z; X, Y) - \delta(\epsilon)]} \end{aligned}$$

where  $D(P||Q)$  is the relative entropy between probability distributions  $P$  and  $Q$ .  $\square$

### C. Proof of Lemma 5.2

We will show that there exists a sequence of codes such that  $\mathbb{P}(\mathcal{D}_{FB,w}) \rightarrow 0$  exponentially, if the given rate constraints are satisfied, which implies the claimed result. The probability of the decoding error event  $\mathcal{D}_{FB,w}$  can be bounded by

$$\begin{aligned}\mathbb{P}(\mathcal{D}_{FB,w}) &= \mathbb{P}(E) \mathbb{P}(\mathcal{D}_{FB,w}|E) + \mathbb{P}(E^c) \mathbb{P}(\mathcal{D}_{FB,w}|E^c) \\ &\leq \mathbb{P}(E) + \mathbb{P}(\mathcal{D}_{FB,w}|E^c) \\ &\leq \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(\mathcal{D}_{FB,w}|E^c)\end{aligned}\quad (62)$$

If we choose the rates of the quantization codebooks such that  $r_i > I(\tilde{V}_j; U_i)$ , for  $(i, j) = (1, 2), (2, 1)$ , by covering lemma [19],  $\mathbb{P}(E_1), \mathbb{P}(E_2) \rightarrow 0$ . Therefore, it is sufficient to show that  $\mathbb{P}(\mathcal{D}_{FB,w}|E^c)$  vanishes if the conditions in the lemma are satisfied.

The decoding error event  $\mathcal{D}_{FB,w}$  can also be expressed as the following union of events.

$$\begin{aligned}\mathcal{D}_{FB,w} &= \bigcup_{m_1 \neq 1} T(m_1, 1, 1) \cup \bigcup_{n_1 \neq 1} T(1, n_1, 1) \\ &\cup \bigcup_{m_2 \neq 1} T(1, 1, m_2) \cup \bigcup_{\substack{m_1 \neq 1 \\ n_1 \neq 1}} T(m_1, n_1, 1) \\ &\cup \bigcup_{\substack{m_1 \neq 1 \\ m_2 \neq 1}} T(m_1, 1, m_2) \cup \bigcup_{\substack{n_1 \neq 1 \\ m_2 \neq 1}} T(1, n_1, m_2) \\ &\cup \bigcup_{\substack{m_1 \neq 1 \\ n_1 \neq 1 \\ m_2 \neq 1}} T(m_1, n_1, m_2) \cup T^c(1, 1, 1)\end{aligned}\quad (63)$$

Using the union bound on (63), probability of decoding error conditioned on quantization success can be bounded by

$$\begin{aligned}\mathbb{P}(\mathcal{D}_{FB,w}|E^c) &= \\ &2^{NR_{1c}} \mathbb{P}(T(m_1, 1, 1)|E^c) + 2^{NR_{1p}} \mathbb{P}(T(1, n_1, 1)|E^c) \\ &+ 2^{NR_{2c}} \mathbb{P}(T(1, 1, m_2)|E^c) + 2^{NR_1} \mathbb{P}(T(m_1, n_1, 1)|E^c) \\ &+ 2^{N(R_{1c}+R_{2c})} \mathbb{P}(T(m_1, 1, m_2)|E^c) \\ &+ 2^{N(R_{1p}+R_{2c})} \mathbb{P}(T(1, n_1, m_2)|E^c) \\ &+ 2^{N(R_1+R_{2c})} \mathbb{P}(T(m_1, n_1, m_2)|E^c) \\ &+ \mathbb{P}(T^c(1, 1, 1)|E^c)\end{aligned}\quad (64)$$

Note that conditioned on successful quantization, the relevant random variables are distributed i.i.d. over time according to the joint distribution

$$\begin{aligned}p(\tilde{x}_{1p}, \tilde{x}_{2e}, \tilde{x}_{2c}, u_1, x_{1e}, x_{1c}, x_{2e}, \tilde{y}_1, y_1) &= \\ p(\tilde{x}_{1p})p(\tilde{x}_{2e})p(\tilde{x}_{2c})p(x_{1e})p(x_{1c})p(x_{2e}) \\ \cdot p(u_1|\tilde{y}_1)p(\tilde{y}_1|\tilde{x}_{2e}, \tilde{x}_{2c}, \tilde{x}_{1p})p(y_1|x_{1e}, x_{1c}, x_{2e})\end{aligned}\quad (65)$$

Next, we bound the error terms one by one. In what follows, joint typicality is sought with respect to the joint distribution in (65). The first term is bounded by  $\mathbb{P}(T^c(1, 1, 1)|E^c) < \epsilon$  by law of large numbers.

The second error term in (64) can be bounded as follows.

$$\mathbb{P}(T(m_1, 1, 1)|E^c) = \mathbb{P}\left(\bigcup_{q_1, q_2, q'_2} T(m_1, 1, 1, q_1, q_2, q'_2) \middle| E^c\right)$$

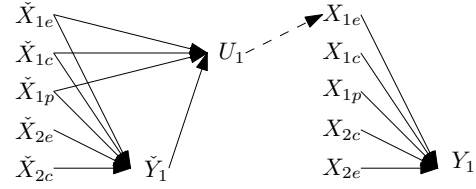


Fig. 12. Markov network showing the dependence of the relevant variables. The variables connected with the dashed arrow are independent in the single-letter form, although in multi-letter form they are not.

$$\begin{aligned}&= \mathbb{P}\left(\bigcup_{\substack{q_1 \neq 1, q_2 \neq 1 \\ q'_2 \neq 1}} T(m_1, 1, 1, q_1, q_2, q'_2) \right. \\ &\cup \bigcup_{\substack{q_2 \neq 1 \\ q'_2 \neq 1}} T(m_1, 1, 1, 1, q_2, q'_2) \cup \bigcup_{\substack{q_1 \neq 1 \\ q'_2 \neq 1}} T(m_1, 1, 1, q_1, 1, q'_2) \\ &\cup \bigcup_{\substack{q_1 \neq 1 \\ q_2 \neq 1}} T(m_1, 1, 1, q_1, q_2, 1) \cup \bigcup_{q_1 \neq 1} T(m_1, 1, 1, q_1, 1, 1) \\ &\cup \bigcup_{q_2 \neq 1} T(m_1, 1, 1, 1, q_2, 1) \cup \bigcup_{q'_2 \neq 1} T(m_1, 1, 1, 1, 1, q'_2) \\ &\left. \cup T(m_1, 1, 1, 1, 1, 1) \middle| E^c\right)\end{aligned}$$

$$\begin{aligned}&\stackrel{(a)}{=} \mathbb{P}\left(\bigcup_{\substack{q_1 \in \mathcal{B}_1(1, q_2) \\ q_2 \in \mathcal{B}_2(b-1), q'_2 \in \mathcal{B}_2(b)}} T(m_1, 1, 1, q_1, q_2, q'_2) \right. \\ &\cup \bigcup_{\substack{q_2 \in \mathcal{B}_2 \\ q'_2 \in \mathcal{B}_2(b)}} T(m_1, 1, 1, 1, q_2, q'_2) \\ &\cup \bigcup_{\substack{q_1 \in \mathcal{B}_1(b) \\ q'_2 \in \mathcal{B}_2(b)}} T(m_1, 1, 1, q_1, 1, q'_2) \\ &\cup \bigcup_{\substack{q_1 \in \mathcal{B}_1(1, q_2) \\ q_2 \in \mathcal{B}_2(b-1)}} T(m_1, 1, 1, q_1, q_2, 1) \\ &\cup \bigcup_{q_1 \in \mathcal{B}_1(b)} T(m_1, 1, 1, q_1, 1, 1) \\ &\cup \bigcup_{q_2 \in \mathcal{B}_2(b-1)} T(m_1, 1, 1, 1, q_2, 1) \\ &\cup \bigcup_{q'_2 \in \mathcal{B}_2(b)} T(m_1, 1, 1, 1, 1, q'_2) \\ &\left. \cup T(m_1, 1, 1, 1, 1, 1) \middle| E^c\right)\end{aligned}$$

$$\stackrel{(b)}{\leq} \mathbb{P}\left(\bigcup_{\substack{q_1 \in \mathcal{B}_1(1, q_2) \\ q_2 \in \mathcal{B}_2(b-1), q'_2 \in \mathcal{B}_2(b)}} T(m_1, 1, 1, q_1, q_2, q'_2) \middle| E^c\right)$$

$$\begin{aligned}
& + \mathbb{P} \left( \bigcup_{\substack{q_2 \in \mathcal{B}_2(b-1) \\ q'_2 \in \mathcal{B}_2(b)}} T(m_1, 1, 1, 1, q_2, q'_2) \middle| E^c \right) \\
& + \mathbb{P} \left( \bigcup_{\substack{q_1 \in \mathcal{B}_1(b) \\ q'_2 \in \mathcal{B}_2(b)}} T(m_1, 1, 1, q_1, 1, q'_2) \middle| E^c \right) \\
& + \mathbb{P} \left( \bigcup_{\substack{q_1 \in \mathcal{B}_1(1, q_2(b-1)) \\ q_2 \in \mathcal{B}_2(b-1)}} T(m_1, 1, 1, q_1, q_2, 1) \middle| E^c \right) \\
& + \mathbb{P} \left( \bigcup_{q_1 \in \mathcal{B}_1(b)} T(m_1, 1, 1, q_1, 1, 1) \middle| E^c \right) \\
& + \mathbb{P} \left( \bigcup_{q_2 \in \mathcal{B}_2(b-1)} T(m_1, 1, 1, 1, q_2, 1) \middle| E^c \right) \\
& + \mathbb{P} \left( \bigcup_{q'_2 \in \mathcal{B}_2(b)} T(m_1, 1, 1, 1, 1, q'_2) \middle| E^c \right) \\
& + \mathbb{P}(T(m_1, 1, 1, 1, 1, 1)|E^c) \\
& \stackrel{(c)}{=} 2^{N(\kappa_1+2\kappa_2)} \mathbb{P}(T(m_1, 1, 1, q_1, q_2, q'_2)|E^c) \\
& \quad + 2^{2N\kappa_2} \mathbb{P}(T(m_1, 1, 1, 1, q_2, q'_2)|E^c) \\
& \quad + 2^{N(\kappa_1+\kappa_2)} \mathbb{P}(T(m_1, 1, 1, q_1, 1, q'_2)|E^c) \\
& \quad + 2^{N(\kappa_1+\kappa_2)} \mathbb{P}(T(m_1, 1, 1, q_1, q_2, 1)|E^c) \\
& \quad + 2^{N\kappa_1} \mathbb{P}(T(m_1, 1, 1, q_1, 1, 1)|E^c) \\
& \quad + 2^{N\kappa_2} \mathbb{P}(T(m_1, 1, 1, 1, q_2, 1)|E^c) \\
& \quad + 2^{N\kappa_2} \mathbb{P}(T(m_1, 1, 1, 1, 1, q'_2)|E^c) \\
& \quad + \mathbb{P}(T(m_1, 1, 1, 1, 1, 1)|E^c) \\
& \stackrel{(d)}{\leq} 2^{N(\kappa_1+2\kappa_2)} \\
& \quad \cdot 2^{-N[I(U_1; \tilde{X}_{2e}) + I(\tilde{X}_{2e}, U_1, X_{1f}, X_{2e}; Y_1, \tilde{Y}_1, \tilde{X}_{2c} | \tilde{X}_1) - \delta(\epsilon)]} \\
& \quad + 2^{2N\kappa_2} 2^{-N[I(X_{1f}, X_{2e}, \tilde{X}_{2e}; \tilde{Y}_1, Y_1 | U_1, X_{1e}, \tilde{X}_{2c}, \tilde{X}_1) - \delta(\epsilon)]} \\
& \quad + 2^{N(\kappa_1+\kappa_2)} \\
& \quad \cdot 2^{-N[I(U_1; \tilde{X}_{2f}) + I(\tilde{X}_{2f}, \tilde{X}_1, U_1, X_{1e}; Y_1, \tilde{Y}_1 | \tilde{X}_{1f}, X_{1c}, X_{2e}) - \delta(\epsilon)]} \\
& \quad + 2^{N(\kappa_1+\kappa_2)} \\
& \quad \cdot 2^{-N[I(\tilde{X}_{2f}, \tilde{X}_1, U_1, X_{1e}, X_{2e}; Y_1, \tilde{Y}_1 | \tilde{X}_{1f}, X_{1c}, \tilde{X}_{2e}) - \delta(\epsilon)]} \\
& \quad + 2^{N\kappa_1} 2^{-N[I(\tilde{X}_1, \tilde{X}_{2f}, U_1, X_{1e}; Y_1, \tilde{Y}_1, \tilde{X}_{2e} | \tilde{X}_{1f}, X_{1c}, X_{2e}) - \delta(\epsilon)]} \\
& \quad + 2^{N\kappa_2} 2^{-N[I(\tilde{X}_1, \tilde{X}_{2f}; U_1, Y_1, \tilde{Y}_1 | X_{1f}, X_{2e}, \tilde{X}_{1f}) - \delta(\epsilon)]} \\
& \quad + 2^{N\kappa_2} 2^{-N[I(\tilde{X}_1, \tilde{X}_{2f}, X_{2e}; Y_1, \tilde{Y}_1, U_1 | \tilde{X}_{1f}, X_{1f}, \tilde{X}_{2e}) - \delta(\epsilon)]} \\
& \quad + 2^{-N[I(\tilde{X}_1, \tilde{X}_{2f}; \tilde{Y}_1, U_1 | Y_1, \tilde{X}_{1f}, \tilde{X}_{2e}, X_{1e}, X_{2e}) - \delta(\epsilon)]} \\
& \stackrel{(e)}{=} 2^{N(\kappa_1+2\kappa_2)} \\
& \quad \cdot 2^{-N[I(U_1; \tilde{X}_{2e}) + I(X_{1f}, X_{2e}; Y_1) + I(\tilde{X}_{2e}; \tilde{Y}_1 | \tilde{X}_1, \tilde{X}_{2c}) - \delta(\epsilon)]} \\
& \quad + 2^{-N[I(X_{1f}, X_{2e}, \tilde{X}_{2e}; \tilde{Y}_1, Y_1 | U_1, X_{1e}, \tilde{X}_{2c}, \tilde{X}_1) - 2\kappa_2 - \delta(\epsilon)]} \\
& \quad + 2^{N(\kappa_1+\kappa_2)} \\
& \quad \cdot 2^{-N[I(U_1; \tilde{X}_{2e}) + I(X_{1f}; Y_1 | X_{2e}) + I(\tilde{X}_{2e}; \tilde{Y}_1 | \tilde{X}_1, \tilde{X}_{2c}) - \delta(\epsilon)]}
\end{aligned}$$

$$\begin{aligned}
& + 2^{-N[I(X_{1f}, X_{2e}; Y_1) + I(U_1; \tilde{Y}_1 | \tilde{X}_1, \tilde{X}_{2f}) - \kappa_1 - \kappa_2 - \delta(\epsilon)]} \\
& + 2^{-N[I(X_{1f}; Y_1 | X_{2e}) + I(U_1; \tilde{Y}_1 | \tilde{X}_1, \tilde{X}_{2f}) - \kappa_1 - \delta(\epsilon)]} \\
& + 2^{-N[I(X_{1f}, \tilde{X}_{2e}; U_1, Y_1, \tilde{Y}_1 | X_{1e}, X_{2e}, \tilde{X}_1, \tilde{X}_{2c}) - \kappa_2 - \delta(\epsilon)]} \\
& + 2^{-N[I(X_{1f}, X_{2e}; Y_1 | X_{1e}) - \kappa_2 - \delta(\epsilon)]} \\
& + 2^{-N[I(X_{1f}; Y_1 | X_{1e}, X_{2e}) - \delta(\epsilon)]} \\
& \stackrel{(f)}{\leq} 8 \cdot 2^{-N[I(X_{1f}; Y_1 | X_{1e}, X_{2e}) - C_1 - \delta(\epsilon)]}
\end{aligned}$$

where

- (a) is since  $T(m_1, 1, 1, q_1, q_2, q'_2)$  is empty set for  $q_1 \notin \mathcal{B}_1(b)$ ,  $q_2 \notin \mathcal{B}_2(b-1)$ , or  $q'_2 \notin \mathcal{B}_2((q_2, m_2)(b))$ , since for random variables  $(X, Y, Z) \sim p(x, y, z)$ ,  $(X, Y, Z) \in \mathcal{T}_\epsilon^{(N)}$  implies  $(X, Y) \in \mathcal{T}_\epsilon^{(N)}$ ,
- (b) follows by union bound,
- (c) follows by Claim B.1, where the upper bound on the size of the  $\mathcal{B}_i$  sets for sufficiently large  $N$  is given by Lemma 5.1,
- (d) follows by packing lemma and Claim B.2,
- (e) is by manipulating the mutual information terms using the dependence structure of the involved variables (see Figure 12),
- (f) is by upper bounding each of the eight terms with the same bound, using chain rule and non-negativity of mutual information.

Next, we bound the term  $\mathbb{P}(T(1, n_1, m_2)|E^c)$ . We apply steps (a)-(d), which are also applicable here, to obtain the following.

$$\begin{aligned}
& \mathbb{P}(T(1, n_1, m_2)|E^c) \\
& \leq 2^{N(\kappa_1+2\kappa_2)} \\
& \quad \cdot 2^{-N[I(U_1; \tilde{X}_{2f}) + I(\tilde{X}_1, \tilde{X}_{2f}, X_{2e}, U_1, X_{1e}; Y_1, \tilde{Y}_1, X_{1c} | \tilde{X}_{1f}) - \delta(\epsilon)]} \\
& \quad + 2^{2N\kappa_2} 2^{-N[I(\tilde{X}_1, \tilde{X}_{2f}, X_{2e}; \tilde{Y}_1, Y_1, U_1 | X_{1f}, \tilde{X}_{1f}) - \delta(\epsilon)]} \\
& \quad + 2^{N(\kappa_1+\kappa_2)} \\
& \quad \cdot 2^{-N[I(U_1; \tilde{X}_{2f}) + I(\tilde{X}_{2f}, \tilde{X}_1, U_1, X_{1e}; Y_1, \tilde{Y}_1 | \tilde{X}_{1f}, X_{1c}, X_{2e}) - \delta(\epsilon)]} \\
& \quad + 2^{N(\kappa_1+\kappa_2)} \\
& \quad \cdot 2^{-N[I(\tilde{X}_{2f}, \tilde{X}_1, U_1, X_{1e}, X_{2e}; Y_1, \tilde{Y}_1 | \tilde{X}_{1f}, X_{1c}, \tilde{X}_{2e}) - \delta(\epsilon)]} \\
& \quad + 2^{N\kappa_1} 2^{-N[I(\tilde{X}_1, \tilde{X}_{2f}, U_1, X_{1e}; Y_1, \tilde{Y}_1, \tilde{X}_{2e} | \tilde{X}_{1f}, X_{1c}, X_{2e}) - \delta(\epsilon)]} \\
& \quad + 2^{N\kappa_2} 2^{-N[I(\tilde{X}_1, \tilde{X}_{2f}; U_1, Y_1, \tilde{Y}_1 | X_{1f}, X_{2e}, \tilde{X}_{1f}) - \delta(\epsilon)]} \\
& \quad + 2^{N\kappa_2} 2^{-N[I(\tilde{X}_1, \tilde{X}_{2f}, X_{2e}; Y_1, \tilde{Y}_1, U_1 | \tilde{X}_{1f}, X_{1f}, \tilde{X}_{2e}) - \delta(\epsilon)]} \\
& \quad + 2^{-N[I(\tilde{X}_1, \tilde{X}_{2f}; \tilde{Y}_1, U_1 | Y_1, \tilde{X}_{1f}, \tilde{X}_{2e}, X_{1e}, X_{2e}) - \delta(\epsilon)]} \\
& \stackrel{(e)}{\leq} 4 \cdot 2^{-N[I(\tilde{X}_{2f}, \tilde{X}_1, U_1, X_{1e}; Y_1, \tilde{Y}_1 | \tilde{X}_{1f}, \tilde{X}_{2e}, X_{1c}, X_{2e}) - \delta(\epsilon)]} \\
& \quad + 4 \cdot 2^{-N[I(\tilde{X}_{2f}, \tilde{X}_1; Y_1, \tilde{Y}_1, U_1 | \tilde{X}_{1f}, \tilde{X}_{2e}, X_{1f}, X_{2e}) - \delta(\epsilon)]} \\
& \stackrel{(f)}{\leq} 4 \cdot 2^{-N[I(X_{2f}, X_1; Y_1 | X_{1c}, X_{2e}) - \delta(\epsilon)]} \\
& \quad + 4 \cdot 2^{-N[I(\tilde{X}_{2f}, \tilde{X}_1; \tilde{Y}_1, U_1 | \tilde{X}_{1f}, \tilde{X}_{2e}) - \delta(\epsilon)]}
\end{aligned}$$

where (e) is by upper bounding the first, third, fourth and fifth terms with the first term in (j), and the rest of the terms with the second; (f) is by rearranging the mutual information terms using chain rule and the fact that the distribution of variables are the same for each block.

In order to bound the term  $\mathbb{P}(T(1, n_1, 1)|E^c)$  in (64), we note that the joint distribution (65) has a similar structure with respect to  $X_{1c}$  and  $\tilde{X}_{1p}$ , with the following mapping between random variables

$$\begin{aligned}\tilde{X}_{1p} &\leftrightarrow X_{1c}, \\ (\tilde{Y}_1, U_1) &\leftrightarrow Y_1, \\ (\tilde{X}_{2e}, \tilde{X}_{2c}) &\leftrightarrow (X_{1e}, X_{2e})\end{aligned}$$

Therefore, one can perform the steps (a)-(f) for the third error term as well, by switching the variables as above, to obtain the following bound

$$\begin{aligned}\mathbb{P}(T(1, n_1, 1)|E^c) &\leq 4 \cdot 2^{-N[I(\tilde{X}_1; \tilde{Y}_1, U_1 | \tilde{X}_{1f}, \tilde{X}_{2f}) - C_1 - \delta(\epsilon)]} \\ &\quad + 4 \cdot 2^{-N[I(\tilde{X}_1; \tilde{Y}_1, U_1 | \tilde{X}_{1c}, \tilde{X}_{2f}) - C_1 - \delta(\epsilon)]} \\ &\leq 8 \cdot 2^{-N[I(\tilde{X}_1; \tilde{Y}_1 | \tilde{X}_{1f}, \tilde{X}_{2f}) - C_1 - \delta(\epsilon)]} \\ &= 8 \cdot 2^{-N[I(X_1; Y_1 | X_{1f}, X_{2f}) - C_1 - \delta(\epsilon)]}\end{aligned}$$

We have dropped the  $U_1$  variable from the mutual information term for the sake of simplicity in evaluating the rate region, since its contribution is small. In the final step, we used the fact that the distribution of variables is the same for each block. We can obtain the following bounds for each error term in a similar way, by exploiting the structure of the joint distribution as done above and noting that the steps (a)-(f) are applicable with an appropriate mapping between the variables.

$$\begin{aligned}\mathbb{P}(T(1, 1, m_2)|E^c) &\leq 8 \cdot 2^{-N[I(\tilde{X}_{2f}; \tilde{Y}_1, U_1 | \tilde{X}_1, \tilde{X}_{2e}) - C_1 - \delta(\epsilon)]} \\ &\leq 8 \cdot 2^{-N[I(\tilde{X}_{2f}; \tilde{Y}_1 | \tilde{X}_1, \tilde{X}_{2e}) - C_1 - \delta(\epsilon)]} \\ &= 8 \cdot 2^{-N[I(X_{2f}; Y_1 | X_1, X_{2e}) - C_1 - \delta(\epsilon)]} \\ \mathbb{P}(T(m_1, n_1, 1)|E^c) &\leq 8 \cdot 2^{-N[I(\tilde{X}_1, X_{1f}; \tilde{Y}_1, Y_1, U_1 | \tilde{X}_{1f}, \tilde{X}_{2f}, X_{1e}, X_{2e}) - C_1 - \delta(\epsilon)]} \\ &\leq 8 \cdot 2^{-N[I(\tilde{X}_1, X_{1f}; \tilde{Y}_1, Y_1 | \tilde{X}_{1f}, \tilde{X}_{2f}, X_{1e}, X_{2e}) - C_1 - \delta(\epsilon)]} \\ &= 8 \cdot 2^{-N[I(\tilde{X}_1; \tilde{Y}_1 | \tilde{X}_{1f}, \tilde{X}_{2f}) + I(X_{1f}; Y_1 | X_{1e}, X_{2e}) - C_1 - \delta(\epsilon)]} \\ &= 8 \cdot 2^{-N[I(X_1; Y_1 | X_{1f}, X_{2f}) + I(X_{1f}; Y_1 | X_{1e}, X_{2e}) - C_1 - \delta(\epsilon)]} \\ \mathbb{P}(T(m_1, 1, m_2)|E^c) &\leq 8 \cdot 2^{-N[I(\tilde{X}_{2f}, X_{1f}; \tilde{Y}_1, Y_1, U_1 | \tilde{X}_1, \tilde{X}_{2e}, X_{1e}, X_{2e}) - C_1 - \delta(\epsilon)]} \\ &\leq 8 \cdot 2^{-N[I(\tilde{X}_{2f}, X_{1f}; \tilde{Y}_1, Y_1 | \tilde{X}_1, \tilde{X}_{2e}, X_{1e}, X_{2e}) - C_1 - \delta(\epsilon)]} \\ &= 8 \cdot 2^{-N[I(\tilde{X}_{2f}; \tilde{Y}_1 | \tilde{X}_1, \tilde{X}_{2e}) + I(X_{1f}; Y_1 | X_{1e}, X_{2e}) - C_1 - \delta(\epsilon)]} \\ &= 8 \cdot 2^{-N[I(X_{2f}; Y_1 | X_1, X_{2e}) + I(X_{1f}; Y_1 | X_{1e}, X_{2e}) - C_1 - \delta(\epsilon)]} \\ \mathbb{P}(T(m_1, n_1, m_2)|E^c) &\leq 8 \cdot 2^{-N[I(\tilde{X}_1, \tilde{X}_{2f}, X_{1f}; \tilde{Y}_1, Y_1, U_1 | \tilde{X}_{1f}, \tilde{X}_{2e}, X_{1e}, X_{2e}) - C_1 - \delta(\epsilon)]} \\ &\leq 8 \cdot 2^{-N[I(\tilde{X}_1, \tilde{X}_{2f}, X_{1f}; \tilde{Y}_1, Y_1 | \tilde{X}_{1f}, \tilde{X}_{2e}, X_{1e}, X_{2e}) - C_1 - \delta(\epsilon)]} \\ &= 8 \cdot 2^{-N[I(\tilde{X}_1, \tilde{X}_{2f}; \tilde{Y}_1 | \tilde{X}_{1f}, \tilde{X}_{2e}) + I(X_{1f}; Y_1 | X_{1e}, X_{2e}) - C_1 - \delta(\epsilon)]} \\ &= 8 \cdot 2^{-N[I(X_1, X_{2f}; Y_1 | X_{1f}, X_{2e}) + I(X_{1f}; Y_1 | X_{1e}, X_{2e}) - C_1 - \delta(\epsilon)]} \\ &= 8 \cdot 2^{-N[I(X_1, X_{2f}; Y_1 | X_{1e}, X_{2e}) - C_1 - \delta(\epsilon)]}\end{aligned}$$

Using these bounds in (64), it is easy to see that if the following are satisfied, then  $\mathbb{P}(\mathcal{D}_{FB,w}|E^c) \rightarrow 0$  as  $N \rightarrow \infty$

(Note that the bounds on  $R_{1p} + R_{1c}$  and  $R_{1c} + R_{2c}$  are redundant, as they can be expressed as a sum of other bounds),

$$R_{1p} < I(X_1; Y_1 | X_{1f}, X_{2f}) - C_1 \quad (66)$$

$$R_{1c} < I(X_{1f}; Y_1 | X_{1e}, X_{2e}) - C_1 \quad (67)$$

$$\begin{aligned}R_{1p} + R_{2c} &< \min \{I(X_{2f}, X_1; Y_1 | X_{1c}, X_{2e}), \\ &\quad I(X_{2f}, X_1; Y_1, U_1 | X_{1f}, X_{2e})\} - C_1 \quad (68)\end{aligned}$$

$$R_1 + R_{2c} < I(X_1, X_{2f}; Y_1 | X_{1e}, X_{2e}) - C_1 \quad (69)$$

The rate constraint on  $R_{1p} + R_{2c}$  provided in the lemma is slightly stricter, which allows us to show the redundancy of some of the bounds obtained later.

#### D. Proof of Lemma 5.4

We will show that there exists a sequence of codes such that  $\mathbb{P}(\mathcal{D}_{FB,s}) \rightarrow 0$  exponentially, if the given rate constraints are satisfied, which implies the claimed result.

Similar to the case of weak interference, choosing the quantization rates such that  $r_i > I(\tilde{V}_j; U_i)$ , for  $(i, j) = (1, 2), (2, 1)$ , probability of decoding error can be bounded by

$$\begin{aligned}\mathbb{P}(\mathcal{D}_{FB,s}|E^c) &= \mathbb{P}(T^c(1, 1, 1)|E^c) + 2^{NR_{1p}}\mathbb{P}(T(1, n_1, 1)|E^c) \\ &\quad + 2^{NR_{2c}}\mathbb{P}(T(1, 1, m_2)|E^c) + 2^{NR_1}\mathbb{P}(T(m_1, n_1, 1)|E^c) \\ &\quad + 2^{N(R_{1p}+R_{2c})}\mathbb{P}(T(1, n_1, m_2)|E^c) \\ &\quad + 2^{N(R_1+R_{2c})}\mathbb{P}(T(m_1, n_1, m_2)|E^c) \quad (70)\end{aligned}$$

Note that conditioned on successful quantization, the relevant random variables are distributed i.i.d. over time according to the joint distribution

$$\begin{aligned}p(\tilde{x}_{1c}, \tilde{x}_{1p}, \tilde{x}_{2e}, x_{2c}, u_2, x_{2e}, x_{1e}, \tilde{y}_1, y_1) \\ = p(\tilde{x}_{1c})p(\tilde{x}_{1p})p(\tilde{x}_{2e})p(x_{1e})p(x_{1c})p(x_{2e}) \\ \cdot p(u_2|\tilde{x}_{1c}, \tilde{x}_{1p})p(\tilde{y}_1|\tilde{x}_{1c}, \tilde{x}_{1c}, \tilde{x}_{2e})p(y_1|x_{1e}, x_{2e}, x_{2c}). \quad (71)\end{aligned}$$

Next, we bound the error terms one by one. In what follows, joint typicality is sought with respect to the joint distribution in (71). The first term is bounded by  $\mathbb{P}(T^c(1, 1, 1)|E^c) < \epsilon$  by law of large numbers.

Now we take the third term, which is bounded as follows.

$$\begin{aligned}\mathbb{P}(T(1, n_1, 1)|E^c) &= \mathbb{P}\left(\bigcup_{q_1, q_2, q'_2} T(1, n_1, 1, q_1, q_2, q'_2)|E^c\right) \\ &= \mathbb{P}\left(\bigcup_{\substack{q_1 \neq 1, q_2 \neq 1 \\ q'_2 \neq 1}} T(1, n_1, 1, q_1, q_2, q'_2) \right. \\ &\quad \cup \bigcup_{\substack{q_2 \neq 1 \\ q'_2 \neq 1}} T(1, n_1, 1, 1, q_2, q'_2) \cup \bigcup_{\substack{q_1 \neq 1 \\ q'_2 \neq 1}} T(1, n_1, 1, q_1, 1, q'_2) \\ &\quad \cup \bigcup_{\substack{q_1 \neq 1 \\ q_2 \neq 1}} T(1, n_1, 1, q_1, q_2, 1) \cup \bigcup_{q_1 \neq 1} T(1, n_1, 1, q_1, 1, 1) \\ &\quad \left. \cup \bigcup_{q_2 \neq 1} T(1, n_1, 1, 1, q_2, 1) \cup \bigcup_{q'_2 \neq 1} T(1, n_1, 1, 1, 1, q'_2)\right)\end{aligned}$$

$$\begin{aligned}
& \cup T(1, n_1, 1, 1, 1, 1) \left| E^c \right) \\
& \stackrel{(a)}{=} \mathbb{P} \left( \bigcup_{\substack{q_1 \in \bar{\mathcal{B}}_1(q_2), q_2 \in \bar{\mathcal{B}}_2(b-1) \\ q'_2 \in \bar{\mathcal{B}}_2(b)}} T(1, n_1, 1, q_1, q_2, q'_2) \right. \\
& \quad \cup \bigcup_{\substack{q_2 \in \bar{\mathcal{B}}_2(b-1) \\ q'_2 \in \bar{\mathcal{B}}_2(b)}} T(1, n_1, 1, 1, q_2, q'_2) \\
& \quad \cup \bigcup_{\substack{q_1 \in \bar{\mathcal{B}}_1(q_2) \\ q'_2 \in \bar{\mathcal{B}}_2(b)}} T(1, n_1, 1, q_1, 1, q'_2) \\
& \quad \cup \bigcup_{\substack{q_1 \in \bar{\mathcal{B}}_1(q_2) \\ q_2 \in \bar{\mathcal{B}}_2(b-1)}} T(1, n_1, 1, q_1, q_2, 1) \\
& \quad \cup \bigcup_{q_1 \in \bar{\mathcal{B}}_1(q_2)} T(1, n_1, 1, q_1, 1, 1) \\
& \quad \cup \bigcup_{q_2 \in \bar{\mathcal{B}}_2(b-1)} T(1, n_1, 1, 1, q_2, 1) \\
& \quad \cup \bigcup_{q'_2 \in \bar{\mathcal{B}}_2(b)} T(1, n_1, 1, 1, 1, q'_2) \\
& \quad \left. \cup T(1, n_1, 1, 1, 1, 1) \left| E^c \right) \right) \\
& \stackrel{(b)}{\leq} \mathbb{P} \left( \bigcup_{\substack{q_1 \in \bar{\mathcal{B}}_1(q_2), q_2 \in \bar{\mathcal{B}}_2(b-1) \\ q'_2 \in \bar{\mathcal{B}}_2(b)}} T(1, n_1, 1, q_1, q_2, q'_2) \left| E^c \right) \right. \\
& \quad + \mathbb{P} \left( \bigcup_{\substack{q_2 \in \bar{\mathcal{B}}_2(b-1) \\ q'_2 \in \bar{\mathcal{B}}_2(b)}} T(1, n_1, 1, 1, q_2, q'_2) \left| E^c \right) \right) \\
& \quad + \mathbb{P} \left( \bigcup_{\substack{q_1 \in \bar{\mathcal{B}}_1(q_2) \\ q'_2 \in \bar{\mathcal{B}}_2(b)}} T(1, n_1, 1, q_1, 1, q'_2) \left| E^c \right) \right) \\
& \quad + \mathbb{P} \left( \bigcup_{\substack{q_1 \in \bar{\mathcal{B}}_1(q_2) \\ q_2 \in \bar{\mathcal{B}}_2(b-1)}} T(1, n_1, 1, q_1, q_2, 1) \left| E^c \right) \right) \\
& \quad + \mathbb{P} \left( \bigcup_{q_1 \in \bar{\mathcal{B}}_1(q_2)} T(1, n_1, 1, q_1, 1, 1) \left| E^c \right) \right) \\
& \quad + \mathbb{P} \left( \bigcup_{q_2 \in \bar{\mathcal{B}}_2(b-1)} T(1, n_1, 1, 1, q_2, 1) \left| E^c \right) \right)
\end{aligned}$$

$$\begin{aligned}
& + \mathbb{P} \left( \bigcup_{q'_2 \in \bar{\mathcal{B}}_2(b)} T(1, n_1, 1, 1, 1, q'_2) \left| E^c \right) \right) \\
& + \mathbb{P} (T(1, n_1, 1, 1, 1, 1) | E^c) \\
& \stackrel{(c)}{\leq} 2^{N(\kappa_1 + 2\kappa_2)} \mathbb{P} (T(1, n_1, 1, q_1, q_2, q'_2) | E^c) \\
& \quad + 2^{2N\kappa_2} \mathbb{P} (T(1, n_1, 1, 1, q_2, q'_2) | E^c) \\
& \quad + 2^{N(\kappa_1 + \kappa_2)} \mathbb{P} (T(1, n_1, 1, q_1, 1, q'_2) | E^c) \\
& \quad + 2^{N(\kappa_1 + \kappa_2)} \mathbb{P} (T(1, n_1, 1, q_1, q_2, 1) | E^c) \\
& \quad + 2^{N\kappa_1} \mathbb{P} (T(1, n_1, 1, \bar{q}_1, 1, 1) | E^c) \\
& \quad + 2^{N\kappa_2} \mathbb{P} (T(1, n_1, 1, 1, \bar{q}_2, 1) | E^c) \\
& \quad + 2^{N\kappa_2} \mathbb{P} (T(1, n_1, 1, 1, 1, \bar{q}'_2) | E^c) \\
& \quad + \mathbb{P} (T(1, n_1, 1, 1, 1, 1) | E^c) \\
& \stackrel{(d)}{\leq} 2^{N(\kappa_1 + 2\kappa_2)} 2^{-NI(U_2; \tilde{X}_{1p} | \tilde{X}_{1e})} \\
& \quad \cdot 2^{-N[I(\tilde{X}_{1p}, \tilde{X}_{2e}, U_2, X_{1e}, X_{2e}; Y_1, \tilde{Y}_1, X_{2c}, \tilde{X}_{1c} | \tilde{X}_{1e}, \tilde{X}_{2c}) - \delta(\epsilon)]} \\
& \quad + 2^{2N\kappa_2} 2^{-NI(U_2; \tilde{X}_{1p})} \\
& \quad \cdot 2^{-N[I(\tilde{X}_{1p}, \tilde{X}_{2e}, U_2, X_{2e}; Y_1, \tilde{Y}_1, X_{1e}, X_{2c}, \tilde{X}_{1c} | \tilde{X}_{1e}, \tilde{X}_{2c}) - \delta(\epsilon)]} \\
& \quad + 2^{N(\kappa_1 + \kappa_2)} 2^{-NI(U_2; \tilde{X}_{1p})} \\
& \quad \cdot 2^{-N[I(\tilde{X}_{1p}, U_2, X_{1e}, X_{2e}; Y_1, \tilde{Y}_1, \tilde{X}_{2e}, X_{2c}, \tilde{X}_{1c} | \tilde{X}_{1e}, \tilde{X}_{2c}) - \delta(\epsilon)]} \\
& \quad + 2^{N(\kappa_1 + \kappa_2)} \\
& \quad \cdot 2^{-N[I(\tilde{X}_{1p}, \tilde{X}_{2e}, X_{1e}; Y_1, \tilde{Y}_1, U_2, X_{2f}, \tilde{X}_{1c} | \tilde{X}_{1e}, \tilde{X}_{2c}) - \delta(\epsilon)]} \\
& \quad + 2^{N\kappa_1} \\
& \quad \cdot 2^{-N[I(\tilde{X}_{1p}, X_{1e}; Y_1, \tilde{Y}_1, \tilde{X}_{2f}, U_2, X_{2f}, \tilde{X}_{1c} | \tilde{X}_{1e}, \tilde{X}_{2c}) - \delta(\epsilon)]} \\
& \quad + 2^{N\kappa_2} 2^{-NI(U_2; \tilde{X}_{1p} | \tilde{X}_{1e})} \\
& \quad \cdot 2^{-N[I(\tilde{X}_1, U_2, X_{2e}; Y_1, \tilde{Y}_1, \tilde{X}_{2e}, X_{2c}, X_{1e}, \tilde{X}_{1c} | \tilde{X}_{1e}, \tilde{X}_{2c}) - \delta(\epsilon)]} \\
& \quad + 2^{N\kappa_2} \\
& \quad \cdot 2^{-N[I(\tilde{X}_{1p}, \tilde{X}_{2e}; Y_1, \tilde{Y}_1, X_{2c}, U_2, X_{1e}, X_{2e}, \tilde{X}_{1c} | \tilde{X}_{1e}, \tilde{X}_{2c}) - \delta(\epsilon)]} \\
& \quad + 2^{-N[I(\tilde{X}_{1p}, Y_1, \tilde{Y}_1, \tilde{X}_{2c}, \tilde{X}_{2e}, U_2, X_{1e}, X_{2e}, \tilde{X}_{1c} | \tilde{X}_{1e}, \tilde{X}_{2c}) - \delta(\epsilon)]} \\
& \stackrel{(e)}{\leq} 2^{N(\kappa_1 + 2\kappa_2)} \\
& \quad \cdot 2^{-N[I(\tilde{X}_{1p}, \tilde{X}_{2e}, X_{1e}, X_{2e}; \tilde{Y}_1, Y_1, U_2 | \tilde{X}_{1f}, X_{2c}, \tilde{X}_{2c}) - \delta(\epsilon)]} \\
& \quad + 2^{2N\kappa_2} \\
& \quad \cdot 2^{-N[I(\tilde{X}_{1p}, \tilde{X}_{2e}, U_2, X_{2e}; Y_1, \tilde{Y}_1 | \tilde{X}_{1f}, X_{1e}, X_{2c}, \tilde{X}_{2c}) - \delta(\epsilon)]} \\
& \quad + 2^{N(\kappa_1 + \kappa_2)} \\
& \quad \cdot 2^{-N[I(\tilde{X}_{1p}, X_{1e}, X_{2e}; \tilde{Y}_1, Y_1, U_2 | \tilde{X}_{1e}, \tilde{X}_{1c}, \tilde{X}_{2f}, X_{2c}) - \delta(\epsilon)]} \\
& \quad + 2^{N(\kappa_1 + \kappa_2)} \\
& \quad \cdot 2^{-N[I(\tilde{X}_{1p}, \tilde{X}_{2e}, X_{1e}; Y_1, \tilde{Y}_1, U_2 | \tilde{X}_{1f}, \tilde{X}_{2c}, X_{2f}) - \delta(\epsilon)]} \\
& \quad + 2^{N\kappa_1} 2^{-N[I(\tilde{X}_{1p}, X_{1e}; Y_1, \tilde{Y}_1, U_2 | \tilde{X}_{1f}, \tilde{X}_{2f}, X_{2f}) - \delta(\epsilon)]} \\
& \quad + 2^{N\kappa_2} 2^{-N[I(\tilde{X}_{1p}, X_{2e}; \tilde{Y}_1, Y_1, U_2 | \tilde{X}_{1f}, \tilde{X}_{2f}, X_{1e}, X_{2c}) - \delta(\epsilon)]} \\
& \quad + 2^{N\kappa_2} 2^{-N[I(\tilde{X}_{1p}, \tilde{X}_{2e}; Y_1, \tilde{Y}_1, U_2 | \tilde{X}_{1f}, \tilde{X}_{2c}, X_{1e}, X_{2f}) - \delta(\epsilon)]} \\
& \quad + 2^{-N[I(\tilde{X}_{1p}; Y_1, \tilde{Y}_1, U_2 | \tilde{X}_{1f}, \tilde{X}_{2f}, X_{1e}, X_{2e}) - \delta(\epsilon)]} \\
& \stackrel{(f)}{\leq} 4 \cdot 2^{-N[I(\tilde{X}_1; \tilde{Y}_1, U_2 | \tilde{X}_{1f}, \tilde{X}_{2f}) - C_1 - \delta(\epsilon)]} \\
& \quad + 4 \cdot 2^{-N[I(\tilde{X}_1, X_{2e}; \tilde{Y}_1, Y_1 | \tilde{X}_{1e}, \tilde{X}_{2f}, X_{1e}, X_{2c}) - C_1 - \delta(\epsilon)]}
\end{aligned}$$

$$\begin{aligned}
&\stackrel{(g)}{=} 4 \cdot 2^{-N[I(X_1;Y_1,U_2|X_{1f},X_{2f})-C_1-\delta(\epsilon)]} \\
&\quad + 4 \cdot 2^{-N[I(X_1,X_{2e};Y_1|X_{1f},X_{2e})-C_1-\delta(\epsilon)]} \\
&\stackrel{(h)}{\leq} 8 \cdot 2^{-N[I(X_1;Y_1|X_{1f},X_{2f})-C_1-\delta(\epsilon)]}
\end{aligned}$$

where

- (a) is since  $T(1, n_1, 1, q_1, q_2, q'_2)$  is empty set for  $q_1 \notin \mathcal{B}_1(b)$ ,  $q_2 \notin \mathcal{B}_2(b-1)$ , or  $q'_2 \notin \mathcal{B}_2((q_2, m_2)(b))$ , since for random variables  $(X, Y, Z) \sim p(x, y, z)$ ,  $(X, Y, Z) \in \mathcal{T}_\epsilon^{(N)}$  implies  $(X, Y) \in \mathcal{T}_\epsilon^{(N)}$ ,
- (b) follows by union bound,
- (c) follows by Claim B.1, where the upper bound on the number of terms is given by Lemma 5.1,
- (d) is by packing lemma, Claim B.2, and the fact that  $X_{1e}^N(b-1)$  is already known at the decoder,
- (e) is by rearranging mutual information terms using chain rule and independence (see Figure 12),
- (f) follows by upper bounding four of the terms with the first expression, the remaining terms with the second expression, and using the definition of  $C_1$ ,
- (g) is because the distribution of variables is the same for all blocks,
- (h) is by upper bounding the two terms with the same expression.

Once again, we use the structure of the joint distribution (71) to show that a similar bounding can be performed for other error terms as follows.

$$\begin{aligned}
&\mathbb{P}(T(1, 1, m_2)|E^c) \\
&\leq 8 \cdot 2^{-N[I(X_{2f};Y_1|X_{1e},X_{2e})-C_1-\delta(\epsilon)]} \\
&\mathbb{P}(T(m_1, n_1, 1)|E^c) \\
&\leq 4 \cdot 2^{-N[I(X_1;Y_1,U_2|X_{1e},X_{2f})-C_1-\delta(\epsilon)]} \\
&\quad + 4 \cdot 2^{-N[I(X_1,X_{2e};Y_1|X_{1e},X_{2e})-C_1-\delta(\epsilon)]} \\
&\mathbb{P}(T(1, n_1, m_2)|E^c) \\
&\leq 8 \cdot 2^{-N[I(\tilde{X}_1,X_{2f};\tilde{Y}_1,Y_1,U_2|\tilde{X}_{1f},\tilde{X}_{2e},X_{1e},X_{2e})-C_1-\delta(\epsilon)]} \\
&= 8 \cdot 2^{-N[I(X_1,X_{2f};Y_1,U_2|X_{1f},X_{2e})-C_1-\delta(\epsilon)]} \\
&\leq 8 \cdot 2^{-N[I(X_1,X_{2f};Y_1|X_{1f},X_{2e})-C_1-\delta(\epsilon)]} \\
&\mathbb{P}(T(m_1, n_1, m_2)|E^c) \\
&\leq 8 \cdot 2^{-N[I(\tilde{X}_1,X_{2f};\tilde{Y}_1,Y_1,U_2|\tilde{X}_{1e},\tilde{X}_{2e},X_{1e},X_{2e})-C_1-\delta(\epsilon)]} \\
&= 8 \cdot 2^{-N[I(X_1,X_{2f};Y_1,U_2|X_{1e},X_{2e})-C_1-\delta(\epsilon)]} \\
&\leq 8 \cdot 2^{-N[I(X_1,X_{2f};Y_1|X_{1e},X_{2e})-C_1-\delta(\epsilon)]}
\end{aligned}$$

Using these bounds in (70), we see that if the conditions in the lemma are satisfied,  $\mathbb{P}(\mathcal{D}_{FBS}|E^c) \rightarrow 0$  as  $N \rightarrow \infty$ .

#### E. Proof of Lemmas 5.3 and 5.5

Extending the notation defined in the first subsection, we define

$$\begin{aligned}
T(m_1, n_1, m_2, q_2) &:= \\
&\{(18) \text{ holds for the indices } (m_1, n_1, m_2, q_2)\}
\end{aligned}$$

We will show that there exists a code such that  $\mathbb{P}(\mathcal{D}_{NFB}) \rightarrow 0$  exponentially, if the given rate constraints are satisfied, which

implies the claimed result. The decoding error event  $\mathcal{D}_{NFB}$  can be expressed as follows.

$$\begin{aligned}
\mathcal{D}_{NFB} &= \left( \bigcap_{q_2} T^c(1, 1, 1, q_2) \right) \cup \bigcup_{n_1 \neq 1} T(1, n_1, 1, 1) \\
&\cup \bigcup_{m_2 \neq 1} T(1, 1, m_2, 1) \cup \bigcup_{\substack{m_1 \neq 1 \\ n_1 \neq 1}} T(m_1, n_1, 1, 1) \\
&\cup \bigcup_{\substack{n_1 \neq 1 \\ m_2 \neq 1}} T(1, n_1, m_2, 1) \cup \bigcup_{\substack{m_2 \neq 1 \\ q_2 \neq 1}} T(1, 1, m_2, q_2) \\
&\cup \bigcup_{\substack{m_1 \neq 1 \\ n_1 \neq 1 \\ m_2 \neq 1}} T(m_1, n_1, m_2, 1) \cup \bigcup_{\substack{n_1 \neq 1 \\ m_2 \neq 1 \\ q_2 \neq 1}} T(1, n_1, m_2, q_2) \\
&\cup \bigcup_{\substack{m_1 \neq 1, n_1 \neq 1 \\ m_2 \neq 1, q_2 \neq 1}} T(m_1, n_1, m_2, q_2)
\end{aligned}$$

Similar to the previous proofs, choosing  $r_i > I(\tilde{V}_j; U_i)$  ensures quantization success with high probability. Then since

$$\mathbb{P}(\mathcal{D}_{NFB}) \leq \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(\mathcal{D}_{NFB}|E^c),$$

it is sufficient to show that  $\mathbb{P}(\mathcal{D}_{NFB}|E^c) \rightarrow 0$ . Using union bound, packing lemma, Lemma 5.1, and Claim B.1, we can upper bound the probability of decoding error conditioned on quantization success by

$$\begin{aligned}
&\mathbb{P}(\mathcal{D}_{NFB}|E^c) \\
&\leq \epsilon_N + 2^{NR_{1p}} 2^{-N[I(X_1;Y_1|X_{1f},X_{2f})-\delta(\epsilon)]} \\
&\quad + 2^{NR_{2c}} 2^{-N[I(X_{2f};Y_1|X_{1e},X_{2e})-\delta(\epsilon)]} \\
&\quad + 2^{NR_1} 2^{-N[I(X_1;Y_1|X_{1e},X_{2f})-\delta(\epsilon)]} \\
&\quad + 2^{N(R_{1p}+R_{2c})} 2^{-N[I(X_1,X_{2f};Y_1|X_{1f},X_{2e})-\delta(\epsilon)]} \\
&\quad + 2^{N(R_{2c}+C'_2)} 2^{-N[I(X_{2f};Y_1|X_{1e})-\delta(\epsilon)]} \\
&\quad + 2^{N(R_1+R_{2c})} 2^{-N[I(X_1,X_{2f};Y_1|X_{1e},X_{2e})-\delta(\epsilon)]} \\
&\quad + 2^{N(R_{1p}+R_{2c}+C'_2)} 2^{-N[I(X_1,X_{2f};Y_1|X_{1f})-\delta(\epsilon)]} \\
&\quad + 2^{N(R_1+R_{2c}+C'_2)} 2^{-N[I(X_1,X_{2f};Y_1|X_{1e})-\delta(\epsilon)]}.
\end{aligned}$$

where  $\epsilon_N \rightarrow 0$  as  $N \rightarrow \infty$ . Note that the conditions in both lemmas are sufficient to ensure  $\mathbb{P}(\mathcal{D}_{NFB}|E^c) \rightarrow 0$  as  $N \rightarrow \infty$ .

### APPENDIX C EVALUATION OF RATE REGIONS

In this section, we consider the set of rate conditions derived in Section V for decodability (*i.e.*, (22)–(25), (27)–(30) for weak interference; (32)–(35), (37)–(39) for strong interference), and obtain an explicit rate region for both linear deterministic and Gaussian models.

#### A. Rate Region for Linear Deterministic Model

Under the input distribution given by (40)–(44), the set of rate constraints for decodability at Rx1 are evaluated as follows:

$$R_{1p} \leq H(Y_1|V_1, V_2) = (n_{11} - n_{21})^+$$

$$\begin{aligned}
R_{2c} &\leq H(Y_1|X_1) = n_{12} \\
R_{1p} + R_{2c} &\leq \min \left\{ H(Y_1, \tilde{V}_2|V_1), H(Y_1) \right\} \\
&= \min \left\{ p_1 (n_{11} - n_{21})^+ \right. \\
&\quad \left. + (1 - p_1) \max \left\{ n_{12}, (n_{11} - n_{21})^+ \right\} + p_1 n_{12}, \right. \\
&\quad \left. \max(n_{11}, n_{12}) \right\} \\
R_1 + R_{2c} &\leq H(Y_1) = \max(n_{11}, n_{12})
\end{aligned}$$

for weak interference ( $n_{12} \leq n_{11}$ ), and

$$\begin{aligned}
R_{1p} &\leq H(Y_1|V_1, V_2) = (n_{11} - n_{21})^+ \\
R_{2c} &\leq H(Y_1|X_1) = n_{12} \\
R_1 &\leq \min \left\{ H(Y_1, \tilde{V}_1|V_2), H(Y_1) \right\} \\
&= \min \left\{ p_2 (n_{11} - n_{21})^+ + (1 - p_2) n_{11} \right. \\
&\quad \left. + p_2 n_{21}, \max(n_{11}, n_{12}) \right\} \\
R_1 + R_{2c} &\leq H(Y_1) = \max(n_{11}, n_{12})
\end{aligned}$$

for strong interference ( $n_{12} > n_{11}$ ). Note that the set of conditions given above can be summarized into the following five inequalities, valid for any interference regime.

$$\begin{aligned}
R_{1p} &\leq (n_{11} - n_{21})^+ \\
R_1 &\leq n_{11} + p_2 (n_{21} - n_{11})^+ \\
R_{2c} &\leq n_{12} \\
R_{1p} + R_{2c} &\leq \max \left\{ n_{12}, (n_{11} - n_{21})^+ \right\} \\
&\quad + p_1 \min \left\{ n_{12}, (n_{11} - n_{21})^+ \right\} \\
R_1 + R_{2c} &\leq \max(n_{11}, n_{12})
\end{aligned}$$

Combining these inequalities with their Rx2 counterparts, and applying Fourier-Motzkin elimination, we arrive at the set of inequalities given in (1)–(6).

### B. Rate Region for Gaussian Model

We consider the input distributions (45)–(49), and the set of input-output relationships given by

$$\begin{aligned}
Y_i &= h_{ii}X_i + h_{ij}X_j + Z_i \\
U_i &= S_i(h_{ij}X_j + Z_i) + Q_i
\end{aligned}$$

for  $(i, j) = (1, 2), (2, 1)$ , where  $Q_i \sim \mathcal{CN}(0, D_i)$ . Choosing  $D_1 = D_2 = \frac{3}{2}$ , and using standard techniques, it is straightforward to evaluate the rate inequalities derived in Section V, and show that the set of rate triples  $(R_{1p}, R_{1c}, R_{2c})$  defined by (72)–(78), for  $(i, j) = (1, 2)$  are contained in the set defined by (22)–(25), (27)–(30) for weak interference, and (32)–(35), (37)–(39) for strong interference. In (72)–(78), we used indicator functions to unify the constraints for weak and strong interference.

In order to find the set of achievable  $(R_1, R_2)$  points, we first note that  $E_i \geq G_i$  and  $C_i \geq G_i$ , and hence the bounds  $E_i$  and  $C_i$  are redundant. Considering the remaining bounds for  $(i, j) = (1, 2), (2, 1)$ , noting that  $F_i \leq A_i + B_i$ , and applying

Fourier-Motzkin elimination, we find that the set of  $(R_1, R_2)$  points that satisfy the following are achievable.

$$R_i < \min \{A_i + B_j, D_i\} \quad (79)$$

$$R_i + R_j < \min \{A_i + G_j, F_i + F_j\} \quad (80)$$

$$2R_i + R_j < A_i + F_j + G_i \quad (81)$$

for  $(i, j) = (1, 2), (2, 1)$ .

## APPENDIX D

### PROOFS OF OUTER BOUNDS (52), (56), (57), AND (58)

In this section, we prove the outer bounds for the linear deterministic channel, based on the ideas presented in Section VI. We first prove four claims that will be useful in the main proof.

#### A. Proof of the Bound (52)

By symmetry, we only focus on the bound on  $R_1$ . By Fano's inequality,

$$\begin{aligned}
N(R_1 - \epsilon_N) &\leq I(W_1; Y_1^N | \underline{S}^N) = I(W_1; Y_1^N, \tilde{V}_1^N, W_2, \underline{S}^N) \\
&= I(W_1; Y_1^N, \tilde{V}_1^N | W_2, \underline{S}^N) \stackrel{(a)}{=} H(Y_1^N, \tilde{V}_1^N | W_2, \underline{S}^N) \\
&= H(Y_1^N | \tilde{V}_1^N, W_2, \underline{S}^N) + H(\tilde{V}_1^N | W_2, \underline{S}^N) \\
&\stackrel{(b)}{=} H(Y_1^N | \tilde{V}_1^N, W_2, X_2^N, \underline{S}^N) + H(\tilde{V}_1^N | W_2, \underline{S}^N) \\
&\leq H(Y_1^N | \tilde{V}_1^N, X_2^N, \underline{S}^N) + H(\tilde{V}_1^N | \underline{S}^N) \\
&\stackrel{(c)}{\leq} n_{11} + p_2 (n_{21} - n_{11})^+
\end{aligned}$$

where (a) follows by the fact that channel is deterministic and hence all variables are completely determined given  $(W_1, W_2, \underline{S}^N)$ ; (b) follows by Claim D.1, and (c) follows by Claim D.5.

#### B. Proof of the Bound (56)

By Fano's inequality,

$$\begin{aligned}
N(R_1 + R_2 - \epsilon_N) &\leq I(W_1; Y_1^N, \underline{S}^N) + I(W_2; Y_2^N, \underline{S}^N) \\
&= I(W_1; Y_1^N | \underline{S}^N) + I(W_2; Y_2^N | \underline{S}^N) \\
&\leq I(W_1; Y_1^N, V_1^N, \tilde{V}_2^N | \underline{S}^N) + I(W_2; Y_2^N, V_2^N, \tilde{V}_1^N | \underline{S}^N) \\
&= H(Y_1^N, V_1^N, \tilde{V}_2^N | \underline{S}^N) + H(Y_2^N, V_2^N, \tilde{V}_1^N | \underline{S}^N) \\
&\quad - H(Y_1^N, V_1^N, \tilde{V}_2^N | W_1, \underline{S}^N) \\
&\quad - H(Y_2^N, V_2^N, \tilde{V}_1^N | W_2, \underline{S}^N) \\
&\stackrel{(a)}{=} H(Y_1^N | V_1^N, \tilde{V}_2^N, \underline{S}^N) + H(Y_2^N | V_2^N, \tilde{V}_1^N, \underline{S}^N) \\
&\quad + H(V_1^N, \tilde{V}_2^N | \underline{S}^N) + H(V_2^N, \tilde{V}_1^N | \underline{S}^N) \\
&\quad - H(V_2^N, \tilde{V}_1^N | W_1, \underline{S}^N) - H(V_1^N, \tilde{V}_2^N | W_2, \underline{S}^N) \\
&= H(Y_1^N | V_1^N, \tilde{V}_2^N, \underline{S}^N) + H(Y_2^N | V_2^N, \tilde{V}_1^N, \underline{S}^N) \\
&\quad + I(W_2; V_1^N, \tilde{V}_2^N | \underline{S}^N) + I(W_1; V_2^N, \tilde{V}_1^N | \underline{S}^N)
\end{aligned}$$



$$R_{ip} < A_i := \log \left( 3 + \frac{\text{SNR}_i}{1 + \text{INR}_j} \right) - \log 3 - C_i \quad (72)$$

$$R_{jc} < B_i := \log (2 + \text{INR}_i) - \log 3 - C_i \quad (73)$$

$$R_i < C_i := \log (3 + \text{SNR}_i + \text{INR}_i) - \log 3 - C_i \quad (74)$$

$$R_i < D_i := \log (3 + \text{SNR}_i) + \mathbb{1}_{\{\text{SNR}_i \leq \text{INR}_i\}} p_j \left[ \log \left( 1 + \frac{\text{INR}_j}{3 + \text{SNR}_i} \right) - \log \frac{5}{3} \right] - \log 3 - C_i \quad (75)$$

$$R_{ip} + R_{jc} < E_i := \log \left( 2 + \text{SNR}_i + \text{INR}_i + \frac{\text{SNR}_i}{1 + \text{INR}_j} \right) - \log 3 - C_i \quad (76)$$

$$R_{ip} + R_{jc} < F_i := \log \left( 2 + \text{INR}_i + \frac{\text{SNR}_i}{1 + \text{INR}_j} \right) + \mathbb{1}_{\{\text{SNR}_i \geq \text{INR}_i\}} p_i \left[ \log \left( \frac{(2 + \text{INR}_i) \left( 3 + \frac{\text{SNR}_i}{1 + \text{INR}_j} \right)}{2 + \frac{\text{SNR}_i}{1 + \text{INR}_j} + \text{INR}_i} \right) - \log 6 \right] \\ - \log 3 - C_i - \mathbb{1}_{\{\text{SNR}_i \geq \text{INR}_i\}} C_i \quad (77)$$

$$R_i + R_{jc} < G_i := \log (2 + \text{SNR}_i + \text{INR}_i) - \log 3 - C_i - \kappa_j \quad (78)$$

$$C_i := p_i + 2p_j, \quad \kappa_j = p_j$$

$$\stackrel{(b)}{\leq} N \max \left\{ n_{12}, (n_{11} - n_{21})^+ \right\} \\ + N \max \left\{ n_{21}, (n_{22} - n_{12})^+ \right\} \\ + N p_1 \min \left\{ n_{12}, (n_{11} - n_{21})^+ \right\} \\ + N p_2 \min \left\{ n_{21}, (n_{22} - n_{12})^+ \right\}$$

where (a) follows by Claim D.2, and (b) follows by Claims D.4 and D.5.

### C. Proof of the Bounds (57) and (58)

By symmetry, it is sufficient to prove (57). To prove this bound, we consider two copies of Rx1, where one of the copies are enhanced as described in Section VI, while the other one is provided with the output of the original channel. The only copy of Rx2 receives the enhanced channel output as well. We would like to prove a sum rate bound for this three-receiver channel. By Fano's inequality,

$$N(2R_1 + R_2 - \epsilon_N) \\ \leq I(W_1; Y_1^N, \underline{S}^N) + I(W_2; Y_2^N, \underline{S}^N) + I(W_1; Y_1^N, \underline{S}^N) \\ = I(W_1; Y_1^N | \underline{S}^N) + I(W_2; Y_2^N | \underline{S}^N) + I(W_1; Y_1^N | \underline{S}^N) \\ \leq I(W_1; Y_1^N | \underline{S}^N) + I(W_2; Y_2^N, V_2^N, \tilde{V}_1^N | \underline{S}^N) \\ + I(W_1; Y_1^N, V_1^N | \underline{S}^N, W_2) \\ \stackrel{(a)}{=} H(Y_1^N | \underline{S}^N) - H(V_2^N, \tilde{V}_1^N | \underline{S}^N, W_1) \\ + H(Y_2^N, V_2^N, \tilde{V}_1^N | \underline{S}^N) - H(Y_2^N, V_2^N, \tilde{V}_1^N | \underline{S}^N, W_2) \\ + H(Y_1^N, V_1^N | \underline{S}^N, W_2) \\ \stackrel{(b)}{=} H(Y_1^N | \underline{S}^N) - H(V_2^N, \tilde{V}_1^N | \underline{S}^N, W_1) \\ + H(V_2^N, \tilde{V}_1^N | \underline{S}^N) + H(Y_2^N | V_2^N, \tilde{V}_1^N) \\ - H(V_1^N | \underline{S}^N, W_2) + H(Y_1^N, V_1^N | \underline{S}^N, W_2) \\ = H(Y_1^N | \underline{S}^N) + I(W_1; V_2^N, \tilde{V}_1^N | \underline{S}^N)$$

$$H(Y_2^N | V_2^N, \tilde{V}_1^N) + H(Y_1^N | \underline{S}^N, W_2, V_1^N)$$

$$\stackrel{(c)}{\leq} \max(n_{11}, n_{12}) + \max \left\{ n_{21}, (n_{22} - n_{12})^+ \right\} \\ + (n_{11} - n_{21})^+ + p_2 \min \left\{ n_{21}, (n_{22} - n_{12})^+ \right\}$$

where (a) follows by Claim D.2, (b) follows by Claim D.3, (c) follows by Claims D.4, D.5 and D.6.

### D. Claims

**Claim D.1.** For  $(i, j) = (1, 2), (2, 1)$ ,

$$X_{i,t} \stackrel{f}{=} (W_i, \tilde{V}_j^{t-1}, \underline{S}^{t-1}) \stackrel{f}{=} (W_i, V_j^{t-1}, \underline{S}^{t-1})$$

*Proof.* We focus on the case  $(i, j) = (1, 2)$  without loss of generality. Note that

$$X_{1,1} \stackrel{f}{=} W_1$$

and by the definition of the channel,

$$X_{1,t} \stackrel{f}{=} (W_1, \tilde{V}_1^{t-1}, \underline{S}^t) \\ \stackrel{(a)}{=} (W_1, \tilde{V}_2^{t-1}, X_1^{t-1} \underline{S}^t),$$

hence the result follows by induction on  $t$ . (a) follows because  $\tilde{Y}_1^{t-1} = S_1^{t-1} \mathbf{H}_{11} X_1^{t-1} + \tilde{V}_2^{t-1}$ .  $\square$

**Claim D.2.** For  $(i, j) = (1, 2), (2, 1)$ ,

$$H(Y_i^N | W_i, \underline{S}^N) = H(Y_j^N, \tilde{V}_i^N | W_i, \underline{S}^N).$$

*Proof.* Let us focus on the case  $(i, j) = (1, 2)$ .

$$H(Y_1^N | W_1, \underline{S}^N) = \sum_{t=1}^N H(Y_{1,t} | W_1, \underline{S}^N, Y_1^{t-1}) \\ \stackrel{(a)}{=} \sum_{t=1}^N H(Y_{1,t} | W_1, \underline{S}^N, Y_1^{t-1}, X_1^t)$$

$$\begin{aligned}
&= \sum_{t=1}^N H(V_{2,t}|W_1, \underline{S}^N, V_2^{t-1}, X_1^t) \\
&\stackrel{(b)}{=} \sum_{t=1}^N H(V_{2,t}|W_1, \underline{S}^N, V_2^{t-1}) \\
&= H(V_2^N, \tilde{V}_1^N|W_1, \underline{S}^N),
\end{aligned}$$

where (a) is by definition, (b) is due to Claim D.1. The other holds similarly.  $\square$

**Claim D.3.** For  $(i, j) = (1, 2), (2, 1)$ ,

$$\begin{aligned}
H(Y_i^N, V_i^N, \tilde{V}_j^N|W_i, \underline{S}^N) &= H(V_j^N, \tilde{V}_i^N|W_i, \underline{S}^N) \\
&= H(V_j^N|W_i, \underline{S}^N)
\end{aligned}$$

*Proof.* Let us focus on the case  $(i, j) = (1, 2)$ .

$$\begin{aligned}
&H(Y_1^N, V_1^N, \tilde{V}_2^N|W_1, \underline{S}^N) \\
&= \sum_{t=1}^N H(Y_{1,t}, V_{1,t}, \tilde{V}_{2,t}|W_1, \underline{S}^N, Y_1^{t-1}, V_1^{t-1}, \tilde{V}_2^{t-1}) \\
&\stackrel{(a)}{=} \sum_{t=1}^N H(Y_{1,t}, V_{1,t}, \tilde{V}_{2,t}|W_1, \underline{S}^N, Y_1^{t-1}, V_1^{t-1}, \tilde{V}_2^{t-1}, X_1^t) \\
&\stackrel{(b)}{=} \sum_{t=1}^N H(V_{2,t}, \tilde{V}_{1,t}|W_1, \underline{S}^N, V_2^{t-1}, \tilde{V}_1^{t-1}, X_1^t) \\
&\stackrel{(c)}{=} \sum_{t=1}^N H(V_{2,t}|W_1, \underline{S}^N, V_2^{t-1}, X_1^t)
\end{aligned}$$

where (a) follows by the fact that  $X_{1,t} \stackrel{f}{=} (W_1, \tilde{Y}_1^N, \underline{S}^N)$ , (b) follows by subtracting  $X_{1,t}$  from  $Y_{1,t}$  and because  $V_{1,t} \stackrel{f}{=} X_{1,t}$  and  $\tilde{V}_{2,t} \stackrel{f}{=} (\underline{S}_t, V_{2,t})$ . Similarly, (c) follows since  $\tilde{V}_{1,t} \stackrel{f}{=} (X_{1,t}, \underline{S}_t)$ .

Now, the two equalities in the claim can be easily obtained from (b) and (c) respectively, by removing  $X_1^t$  from the conditioning by virtue of Claim D.1, and using chain rule.  $\square$

**Claim D.4.** For  $(i, j) = (1, 2), (2, 1)$ ,

$$I(W_i; V_j^N, \tilde{V}_i^N|\underline{S}^N) \leq N p_j n_{ji}.$$

*Proof.* Let us focus on the case  $(i, j) = (1, 2)$ .

$$\begin{aligned}
&I(W_1; V_2^N, \tilde{V}_1^N|\underline{S}^N) \\
&\stackrel{(a)}{\leq} I(W_1; W_2, \tilde{V}_1^N|\underline{S}^N) = I(W_1; \tilde{V}_1^N|\underline{S}^N, W_2) \\
&= H(\tilde{V}_1^N|\underline{S}^N, W_2) \leq H(\tilde{V}_1^N|S_2^N) \\
&= \mathbb{E}_{S_2^N} [H((s_2 V_1)^N) | S_2^N = s_2^N] \\
&\leq \mathbb{E}_{S_2^N} \left[ \sum_{t=1}^N H(s_{2,t} V_{1,t}) \middle| S_2^N = s_2^N \right] \\
&\leq \mathbb{E}_{S_2^N} [N_1(s_2^N) n_{21} | S_2^N = s_2^N] \\
&= N p_2 n_{21}.
\end{aligned}$$

Here  $N_1(\cdot)$  denotes the number of 1's in the sequence. (a) follows because  $V_2^N \stackrel{f}{=} (W_2, \tilde{V}_1^N, \underline{S}^N)$ .  $\square$

**Claim D.5.** For  $(i, j) = (1, 2), (2, 1)$ ,

$$\begin{aligned}
&N^{-1} H(Y_i^N|V_i^N, \tilde{V}_j^N, \underline{S}^N) \\
&\leq p_i(n_{ii} - n_{ji})^+ + (1 - p_i) \max\{n_{ij}, (n_{ii} - n_{ji})^+\}, \\
&N^{-1} H(Y_i^N|V_j^N, \tilde{V}_i^N, \underline{S}^N) \\
&\leq p_j(n_{ii} - n_{ji})^+ + (1 - p_j)n_{ii},
\end{aligned}$$

*Proof.* Let us focus on the case  $(i, j) = (1, 2)$ .

$$\begin{aligned}
&H(Y_1^N|V_1^N, \tilde{V}_2^N, \underline{S}^N) \leq H(Y_1^N|V_1^N, \tilde{V}_2^N, S_1^N) \\
&= \mathbb{E}_{S_1^N} [H(Y_1^N|V_1^N, (s_1 V_2)^N) | S_1^N = s_1^N] \\
&\leq \mathbb{E}_{S_1^N} \left[ \sum_{t=1}^N H(Y_{1,t}|V_{1,t}, s_{1,t} V_{2,t}) \middle| S_1^N = s_1^N \right] \\
&\leq \mathbb{E}_{S_1^N} \left[ \frac{N_1(s_1^N)}{+N_0(s_1^N)} \max\{n_{12}, (n_{11} - n_{21})^+\} \middle| S_1^N = s_1^N \right] \\
&= N p_1(n_{11} - n_{21})^+ + N(1 - p_1) \max\{n_{12}, (n_{11} - n_{21})^+\}
\end{aligned}$$

Here  $N_1(\cdot)$  and  $N_0(\cdot)$  denote the number of 1's and 0's respectively in the sequence.

For the second inequality,

$$\begin{aligned}
&H(Y_1^N|V_2^N, \tilde{V}_1^N, \underline{S}^N) \leq H(Y_1^N|V_2^N, \tilde{V}_1^N, S_2^N) \\
&= \mathbb{E}_{S_2^N} [H(Y_1^N|V_2^N, (s_2 V_1)^N) | S_2^N = s_2^N] \\
&\leq \mathbb{E}_{S_2^N} \left[ \sum_{t=1}^N H(Y_{1,t}|V_{2,t}, s_{2,t} V_{1,t}) \middle| S_2^N = s_2^N \right] \\
&\leq \mathbb{E}_{S_2^N} [N_1(s_2^N) (n_{11} - n_{21})^+ + N_0(s_2^N) n_{11} | S_2^N = s_2^N] \\
&= N p_2(n_{11} - n_{21})^+ + N(1 - p_2)n_{11}.
\end{aligned}$$

The case  $(i, j) = (2, 1)$  follows similarly.  $\square$

**Claim D.6.** For  $(i, j) = (1, 2), (2, 1)$ ,

$$H(Y_i^N|\underline{S}^N, W_j, V_i^N) \leq N(n_{ii} - n_{ji})^+$$

*Proof.* We focus on  $(i, j) = (1, 2)$  without loss of generality.

$$\begin{aligned}
&H(Y_i^N|\underline{S}^N, W_j, V_i^N) \stackrel{(a)}{=} H(Y_i^N|\underline{S}^N, W_j, V_i^N, V_j^N) \\
&\leq H(Y_i^N|\underline{S}^N, V_i^N, V_j^N) \leq N(n_{11} - n_{21})^+
\end{aligned}$$

where (a) follows because  $V_j^N \stackrel{f}{=} X_j^N \stackrel{f}{=} (W_j, V_i^N, \underline{S}^N)$  by Claim D.1.  $\square$

## APPENDIX E

### PROOFS OF OUTER BOUNDS (53), (59), (60), AND (61)

In this section, we prove an outer bound region for the enhanced channel defined in Section VI.

### A. Notation

We define

$$\check{V}_i = \begin{cases} \bar{V}_i, & \text{if } S_j = 0 \\ Y_{ji}, & \text{if } S_j = 1 \end{cases}$$

for  $(i, j) = (1, 2), (2, 1)$ , where  $\bar{V}_i = Y_{ji} + Z_{jj} = h_{ji}X_i + \bar{Z}_j$ , and

$$M_i = |\{t : S_{i,t} = 1\}|, \\ L_i = N - M_i.$$

For any random vector  $E^N$ , we define

$$E^{(t)} = \{E_{t'}\}_{t': S_{i,t'}=1, t' \leq t} \\ E^{[t]} = \{E_{t'}\}_{t': S_{i,t'}=0, t' \leq t} \\ E_{i,(t)} = \begin{cases} \emptyset, & \text{if } S_i = 0 \\ E_{i,t}, & \text{if } S_i = 1 \end{cases} \\ E_{i,[t]} = \begin{cases} E_{i,t}, & \text{if } S_i = 0 \\ \emptyset, & \text{if } S_i = 1 \end{cases}$$

for  $i = 1$  or  $2$ . Note that in vector form, this notation omits any reference to user index  $i$  for the sake of brevity. That is, although it is not clear whether  $E^{(t)}$  is defined with respect to  $S_1$  or  $S_2$ , in the proof this will be clear from the context. For instance,  $Y_1^{(t)}$  and  $V_2^{(t)}$  are defined with respect to  $S_1$ , since these variables refer to signals that pass through the feedback channel controlled by  $S_1$ . The partial average power for Tx $i$ ,  $P_i^{(jk)}$ , is a random variable defined as

$$P_i^{(j0)} = \frac{1}{L_i} \sum_{t: S_{j,t}=0} P_{i,t} \\ P_i^{(j1)} = \frac{1}{M_i} \sum_{t: S_{j,t}=1} P_{i,t}$$

for  $j = 1, 2$ , where  $P_{i,t}$  is the power used by Tx $i$  at time slot  $t$ .

Finally, we define  $h_S(\cdot) := h(\cdot | \underline{S}^N = S^N)$  for convenience, where  $h(\cdot)$  denotes differential entropy, and  $S^N$  is a particular realization of  $\underline{S}^N$ . Similarly, we define  $I_S(\cdot; \cdot) := I_S(\cdot; \cdot | \underline{S}^N = S^N)$ .

### B. Proof of the Bound (53)

We focus on the case  $(i, j) = (1, 2)$ . By Fano's inequality,

$$\begin{aligned} N(R_1 - \epsilon_N) &\leq I(W_1; Y_1^N, \underline{S}^N) \\ &\leq I(W_1; Y_1^N, \tilde{V}_1^N, W_2, \underline{S}^N) \\ &\leq I(W_1; Y_1^N, \tilde{V}_1^N, \underline{S}^N | W_2) \\ &\leq \sum_{t=1}^N I(W_1; Y_{1,t}, \tilde{V}_{1,t}, \underline{S}_t | W_2, Y_1^{t-1}, \tilde{V}_1^{t-1}, \underline{S}^{t-1}) \\ &= \sum_{t=1}^N I(W_1; Y_{1,t}, \tilde{V}_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^{t-1}, \underline{S}^t) \\ &\stackrel{(a)}{=} \sum_{t=1}^N I(W_1; Y_{1,t}, \tilde{V}_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^{t-1}, \underline{S}^t, X_{2,t}) \\ &= \sum_{t=1}^N I(W_1; Y_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^t, \underline{S}^t, X_{2,t}) \end{aligned}$$

$$\begin{aligned} &+ I(W_1; \tilde{V}_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^{t-1}, \underline{S}^t, X_{2,t}) \\ &= \sum_{t=1}^N h(Y_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^t, \underline{S}^t, X_{2,t}) \\ &\quad - h(Y_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^t, \underline{S}^t, X_{2,t}, W_1) \\ &\quad + I(W_1; \tilde{V}_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^{t-1}, \underline{S}^t, X_{2,t}) \\ &\stackrel{(b)}{=} \sum_{t=1}^N h(Y_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^t, \underline{S}^t, X_{2,t}) \\ &\quad - h(Y_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^t, \underline{S}^t, X_{2,t}, W_1, X_{1,t}) \\ &\quad + \sum_{t=1}^N I(W_1; \tilde{V}_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^{t-1}, \underline{S}^t, X_{2,t}) \\ &= \sum_{t=1}^N h(Y_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^t, \underline{S}^t, X_{2,t}) \\ &\quad - h(Z_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^t, \underline{S}^t, X_{2,t}, W_1, X_{1,t}) \\ &\quad + I(W_1; \tilde{V}_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^{t-1}, \underline{S}^t, X_{2,t}) \\ &\stackrel{(c)}{=} \sum_{t=1}^N h(Y_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^t, \underline{S}^t, X_{2,t}) - h(Z_{1,t}) \\ &\quad + I(W_1; \tilde{V}_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^{t-1}, \underline{S}^t, X_{2,t}) \\ &\leq \sum_{t=1}^N h(Y_{1,t} | \tilde{V}_{1,t}, S_{2,t}, X_{2,t}) - h(Z_{1,t}) \\ &\quad + I(W_1; \tilde{V}_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^{t-1}, \underline{S}^t, X_{2,t}) \\ &\leq \sum_{t=1}^N h(Y_{1,t} | \tilde{V}_{1,t}, S_{2,t}, X_{2,t}) - h(Z_{1,t}) \\ &\quad + I(W_1, X_{1,t}; \tilde{V}_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^{t-1}, \underline{S}^t, X_{2,t}) \\ &= \sum_{t=1}^N h(Y_{1,t} | \tilde{V}_{1,t}, S_{2,t}, X_{2,t}) - h(Z_{1,t}) \\ &\quad + I(X_{1,t}; \tilde{V}_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^{t-1}, \underline{S}^t, X_{2,t}) \\ &\quad + I(W_1; \tilde{V}_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^{t-1}, \underline{S}^t, X_{2,t}, X_{1,t}) \\ &\stackrel{(d)}{=} \sum_{t=1}^N h(Y_{1,t} | \tilde{V}_{1,t}, S_{2,t}, X_{2,t}) - h(Z_{1,t}) \\ &\quad + I(X_{1,t}; \tilde{V}_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^{t-1}, \underline{S}^t, X_{2,t}) \\ &= \sum_{t=1}^N h(Y_{1,t} | \tilde{V}_{1,t}, S_{2,t}, X_{2,t}) - h(Z_{1,t}) \\ &\quad + h(\tilde{V}_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^{t-1}, \underline{S}^t, X_{2,t}) \\ &\quad - h(\tilde{V}_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^{t-1}, \underline{S}^t, X_{2,t}, X_{1,t}) \\ &\leq \sum_{t=1}^N h(Y_{1,t} | \tilde{V}_{1,t}, S_{2,t}, X_{2,t}) - h(Z_{1,t}) \\ &\quad + h(\tilde{V}_{1,t} | S_{2,t}) - h(\tilde{V}_{1,t} | W_2, Y_1^{t-1}, \tilde{V}_1^{t-1}, \underline{S}^t, X_{2,t}, X_{1,t}) \\ &\stackrel{(e)}{=} \sum_{t=1}^N h(Y_{1,t} | \tilde{V}_{1,t}, S_{2,t}, X_{2,t}) - h(Z_{1,t}) \\ &\quad + h(\tilde{V}_{1,t} | \underline{S}_t) - h(\tilde{V}_{1,t} | S_{2,t}, X_{1,t}) \end{aligned}$$

$$\begin{aligned}
&= \sum_{t=1}^N h(Y_{1,t}|\tilde{V}_{1,t}, S_{2,t}, X_{2,t}) - h(Z_{1,t}) \\
&\quad + I(X_{1,t}; \tilde{V}_{1,t}|S_{2,t}) \\
&\stackrel{(f)}{=} p_2 \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) + (1 - p_2) \log(1 + \text{SNR}_1) \\
&\quad + p_2 \log(1 + \text{INR}_2) \\
&= \log(1 + \text{SNR}_1) + p_2 \log \left( 1 + \frac{\text{INR}_2}{1 + \text{SNR}_1} \right)
\end{aligned}$$

where

- (a) is due to Lemma E.1
- (b) is because  $X_{1,t} \stackrel{f}{=} (\underline{S}^{t-1}, W_1, \tilde{Y}_1^{t-1}) \stackrel{f}{=} (\underline{S}^{t-1}, W_1, Y_1^{t-1})$ ,
- (c) is because  $Z_{1,t}$  is independent from all past signals and messages,
- (d) is because  $W_1 - X_{1,t} - \tilde{V}_{1,t}$  is a Markov chain, hence conditioned on  $X_{1,t}$ ,  $\tilde{V}_{1,t}$  is independent from  $W_1$  and all the other past signals,
- (e) is because given  $(S_{2,t}, X_{1,t})$ ,  $\tilde{V}_{1,t}$  is independent from all the other variables in the conditioning,
- (f) is due to Lemma E.2

### C. Proof of Bound (59)

In this section, we exclusively focus on the enhanced channel defined in Section VI. By Fano's inequality.

$$\begin{aligned}
&N(R_1 + R_2 - \epsilon_N) \\
&\leq I(W_1; \check{Y}_1^N, \underline{S}^N) + I(W_2; \check{Y}_2^N, \underline{S}^N) \\
&= I(W_1; \check{Y}_1^N | \underline{S}^N) + I(W_2; \check{Y}_2^N | \underline{S}^N) \\
&\leq I(W_1; \check{Y}_1^N, \check{V}_1^N | \underline{S}^N) + I(W_2; \check{Y}_2^N, \check{V}_2^N | \underline{S}^N) \\
&= h(\check{Y}_1^N | \underline{S}^N, \check{V}_1^N) + h(\check{Y}_2^N | \underline{S}^N, \check{V}_2^N) \quad (82) \\
&\quad + h(\check{V}_1^N | \underline{S}^N) + h(\check{V}_2^N | \underline{S}^N) \quad (83) \\
&\quad - h(\check{Y}_1^N, \check{V}_1^N | \underline{S}^N, W_1) - h(\check{Y}_2^N, \check{V}_2^N | \underline{S}^N, W_2) \quad (84)
\end{aligned}$$

Let us take one term from (82).

$$\begin{aligned}
&h(\check{Y}_1^N | \underline{S}^N, \check{V}_1^N) = \mathbb{E}_{S^N} \left[ h_S \left( \check{Y}_1^N | \check{V}_1^N \right) \right] \\
&\stackrel{(a)}{=} \mathbb{E}_{S^N} \left[ h_S \left( \check{Y}_1^{L_1}, Y_{11}^{M_1}, Y_{12}^{M_1} | \check{V}_1^N \right) \right] \\
&\leq \mathbb{E}_{S^N} \left[ h_S \left( \check{Y}_1^{L_1} | \check{V}_1^N \right) \right] + \mathbb{E}_{S^N} \left[ h_S \left( Y_{11}^{M_1} | \check{V}_1^N \right) \right] \\
&\quad + \mathbb{E}_{S^N} \left[ h_S \left( Y_{12}^{M_1} | \check{V}_1^N \right) \right]
\end{aligned}$$

where (a) follows by (with a slight abuse of notation) decomposing  $\check{Y}_1^N$  into  $(Y_{11}^{M_1}, Y_{12}^{M_1})$  for time slots where  $S_{1,t} = 1$ , and into  $\check{Y}_1^{L_1}$  for time slots for which  $S_{1,t} = 0$ .

The other term in (82) can be bounded similarly. Let us take one term from (84).

$$\begin{aligned}
&-h(\check{Y}_1^N, \check{V}_1^N | \underline{S}^N, W_1) \\
&= -\mathbb{E}_{S^N} \left[ h_S \left( \check{Y}_1^N, \check{V}_1^N | W_1 \right) \right] \\
&\stackrel{(b)}{=} -\mathbb{E}_{S^N} \left[ h_S \left( \check{V}_2^N, Z_{11}^{M_1}, Z_{22}^{M_2} | W_1 \right) \right]
\end{aligned}$$

$$\begin{aligned}
&\stackrel{(c)}{=} -\sum_{t=1}^N \mathbb{E}_{S^N} \left[ h \left( \check{V}_{2,t}, Z_{11,(t)}, Z_{2,[t]}, Z_{21,(t)} \right. \right. \\
&\quad \left. \left. | W_1, \check{V}_2^{t-1}, Z_{11}^{(t-1)}, Z_2^{[t-1]}, Z_{21}^{(t-1)} \right) \right] \\
&\stackrel{(d)}{=} -\mathbb{E} \left[ \sum_{t=1}^N h_S \left( \check{V}_{2,t} | W_1, \check{V}_2^{t-1}, Z_{11}^{(t-1)}, Z_2^{[t-1]}, Z_{21}^{(t-1)} \right) \right. \\
&\quad + \sum_{t:S_{1,t}=1} h_S \left( Z_{11,t} | W_1, \check{V}_2^{t-1}, Z_{11}^{(t-1)}, Z_2^{[t-1]}, Z_{21}^{(t-1)} \right) \\
&\quad + \sum_{t:S_{2,t}=0} h_S \left( Z_{2,t} | W_1, \check{V}_2^{t-1}, Z_{11}^{(t-1)}, Z_2^{[t-1]}, Z_{21}^{(t-1)} \right) \\
&\quad \left. + \sum_{t:S_{2,t}=1} h_S \left( Z_{21,t} | W_1, \check{V}_2^{t-1}, Z_{11}^{(t-1)}, Z_2^{[t-1]}, Z_{21}^{(t-1)} \right) \right] \\
&\stackrel{(e)}{=} -\mathbb{E}_{S^N} \left[ \sum_{t=1}^N h_S \left( \check{V}_{2,t} | W_1, \check{V}_2^{t-1}, Z_{11}^{(t-1)}, Z_2^{[t-1]}, Z_{21}^{(t-1)} \right) \right] \\
&\quad - \mathbb{E}_{S^N} [L_2 h(Z_2) + M_2 h(Z_{21}) + M_1 h(Z_{11})] \\
&= -\mathbb{E}_{S^N} \left[ \sum_{t=1}^N h_S \left( \check{V}_{2,t} | W_1, \check{V}_2^{t-1}, Z_{11}^{(t-1)}, Z_2^{[t-1]}, Z_{21}^{(t-1)} \right) \right] \\
&\quad - N(1 - p_2)h(Z_2) - Np_2h(Z_{21}) - Np_1h(Z_{11})
\end{aligned}$$

where (b) follows by Lemma E.6, (c) follows by chain rule, (d) follows by the fact that for a given time slot  $t$ , the involved noise terms are independent from each other and from  $\check{V}_{2,t}$ ; and (e) is because the signals up to time  $t-1$  are independent from the noise at time  $t$ , and because noise processes are i.i.d.

The other term in (84) can be bounded similarly.

Putting everything together, we have

$$\begin{aligned}
&N(R_1 + R_2 - \epsilon_N) \\
&\leq \mathbb{E}_{S^N} \left[ h_S \left( \check{Y}_1^{L_1} | \check{V}_1^N \right) \right] + \mathbb{E}_{S^N} \left[ h_S \left( Y_{11}^{M_1} | \check{V}_1^N \right) \right] \\
&\quad + \mathbb{E}_{S^N} \left[ h_S \left( \check{Y}_2^{L_2} | \check{V}_2^N \right) \right] + \mathbb{E}_{S^N} \left[ h_S \left( Y_{22}^{M_2} | \check{V}_2^N \right) \right] \\
&\quad + \mathbb{E}_{S^N} \left[ h_S \left( \check{V}_1^N, Y_{12}^{M_1} \right) \right] \quad (85) \\
&\quad + \mathbb{E}_{S^N} \left[ h_S \left( \check{V}_2^N, Y_{21}^{M_2} \right) \right] \quad (86)
\end{aligned}$$

$$\begin{aligned}
&- \mathbb{E}_{S^N} \left[ \sum_{t=1}^N h_S \left( \check{V}_{2,t} | W_1, \check{V}_2^{t-1}, Z_{11}^{(t-1)}, Z_2^{[t-1]}, Z_{21}^{(t-1)} \right) \right] \quad (87)
\end{aligned}$$

$$\begin{aligned}
&- \mathbb{E}_{S^N} \left[ \sum_{t=1}^N h_S \left( \check{V}_{1,t} | W_2, \check{V}_1^{t-1}, Z_{22}^{(t-1)}, Z_1^{[t-1]}, Z_{12}^{(t-1)} \right) \right] \quad (88)
\end{aligned}$$

$$\begin{aligned}
&- N(1 - p_2)h(Z_2) - Np_2h(Z_{21}) - Np_1h(Z_{11}) \\
&- N(1 - p_1)h(Z_1) - Np_1h(Z_{12}) - Np_2h(Z_{22})
\end{aligned}$$

Let us combine (85) and (88).

$$\begin{aligned}
&(85) + (88) = \mathbb{E}_{S^N} \left[ h_S \left( \check{V}_1^N, Y_{12}^{M_1} \right) \right. \\
&\quad \left. - \sum_{t=1}^N h_S \left( \check{V}_{1,t} | W_2, \check{V}_1^{t-1}, Z_{22}^{(t-1)}, Z_1^{[t-1]}, Z_{12}^{(t-1)} \right) \right] \\
&= \mathbb{E}_{S^N} \left[ \sum_{t=1}^N h_S \left( \check{V}_{1,t} | \check{V}_1^{t-1}, Y_{12}^{(t-1)} \right) \right]
\end{aligned}$$

$$\begin{aligned}
& -h_S(\check{V}_{1,t}|W_2, \check{V}_1^{t-1}, Z_{22}^{(t-1)}, Z_1^{[t-1]} Z_{12}^{(t-1)}) \\
& + \sum_{t:S_{1,t}=1} h_S(Y_{12,t}|\check{V}_1^t, Y_{12}^{(t-1)}) \Big] \\
& \leq \mathbb{E}_{S^N} \left[ \sum_{t:S_{1,t}=1} h_S(Y_{12,t}) \right. \\
& \quad \left. \sum_{t=1}^N I_S(\check{V}_{1,t}; W_2, Z_{22}^{(t-1)}, Z_1^{[t-1]} Z_{12}^{(t-1)} | \check{V}_1^{t-1}, Y_{12}^{(t-1)}) \right]
\end{aligned}$$

Similarly, we combine (86) with (87) to obtain the same expression with user indices swapped. Plugging these back, we get

$$\begin{aligned}
& N(R_1 + R_2 - \epsilon_N) \\
& \leq \mathbb{E}_{S^N} \left[ h_S(\bar{Y}_1^{L_1} | \bar{V}_1^N) \right] + \mathbb{E}_{S^N} \left[ h_S(\bar{Y}_{11}^{M_1} | \bar{V}_1^N) \right] \\
& + \mathbb{E}_{S^N} \left[ h_S(\bar{Y}_2^{L_2} | \bar{V}_2^N) \right] + h_S(\bar{Y}_{22}^{M_2} | \bar{V}_2^N) \\
& + \sum_{t=1}^N I_S(\check{V}_{1,t}; W_2, Z_{22}^{(t-1)}, Z_1^{[t-1]} Z_{12}^{(t-1)} | \check{V}_1^{t-1}, Y_{12}^{(t-1)}) \\
& + \sum_{t=1}^N I_S(\check{V}_{2,t}; W_1, Z_{11}^{(t-1)}, Z_2^{[t-1]} Z_{21}^{(t-1)} | \check{V}_2^{t-1}, Y_{21}^{(t-1)}) \Big] \\
& + \mathbb{E}_{S^N} \left[ \sum_{t:S_{1,t}=1} h_S(Y_{12,t}) \right] + \mathbb{E}_{S^N} \left[ \sum_{t:S_{2,t}=1} h_S(Y_{21,t}) \right] \\
& - N(1-p_2)h(Z_2) - Np_2h(Z_{21}) - Np_1h(Z_{11}) \\
& - N(1-p_1)h(Z_1) - Np_1h(Z_{12}) - Np_2h(Z_{22})
\end{aligned}$$

We use Lemmas E.3, E.4, E.7 and E.8 to bound each of these terms, and use the fact that noise distribution is Gaussian to obtain the desired bound.

#### D. Proof of the Bounds (60) and (61)

We exclusively focus on the enhanced channel, defined in Section VI. By symmetry, it is sufficient to prove (60). In addition to the enhanced interference channel in the case of sum rate bound, we consider an additional copy of Receiver 1, who always receives  $\bar{Y}_1$  (i.e., as in the original channel). The feedback signal of Tx1 is still given by  $S_1 \cdot \bar{Y}_1$ , i.e., the same as the original channel. We would like to prove a sum rate upper bound on this new channel.

By Fano's inequality,

$$\begin{aligned}
& N(2R_1 + R_2 - \epsilon_N) \\
& \leq I(W_1; \bar{Y}_1^N, \underline{S}^N) + I(W_2; \check{Y}_2^N, \underline{S}^N) + I(W_1; \check{Y}_1^N, \underline{S}^N) \\
& = I(W_1; \bar{Y}_1^N | \underline{S}^N) + I(W_2; \check{Y}_2^N | \underline{S}^N) + I(W_1; \check{Y}_1^N | \underline{S}^N) \\
& = I(W_1; \bar{Y}_1^N | \underline{S}^N) + I(W_2; \check{Y}_2^N, \bar{V}_2^N | \underline{S}^N) \\
& + I(W_1; \check{Y}_1^N, \check{V}_1^N | W_2, \underline{S}^N) \tag{89}
\end{aligned}$$

The first mutual information term in (89) can be bounded as follows

$$I(W_1; \bar{Y}_1^N | \underline{S}^N) = h(\bar{Y}_1^N | \underline{S}^N) - h(\bar{Y}_1^N | W_1, \underline{S}^N)$$

$$\begin{aligned}
& = h(\bar{Y}_1^N | \underline{S}^N) - \sum_{t=1}^N h(\bar{Y}_{1,t} | W_1, \bar{Y}_1^{t-1}, \underline{S}^N) \\
& \stackrel{(a)}{=} h(\bar{Y}_1^N | \underline{S}^N) - \sum_{t=1}^N h(\bar{V}_{2,t} | W_1, \bar{V}_2^{t-1}, \underline{S}^N) \\
& \leq \sum_{t=1}^N h(\bar{Y}_{1,t}) - h(\bar{V}_{2,t} | W_1, \bar{V}_2^{t-1}, \underline{S}^N)
\end{aligned}$$

where (a) follows by the fact that  $X_{1,t} \stackrel{f}{=} (W_1, \bar{Y}_1^{t-1}, \underline{S}^{t-1})$  and by subtracting  $X_{1,t}$  from  $\bar{Y}_{1,t}$ .

Let us consider the second mutual information term from (89).

$$\begin{aligned}
& I(W_2; \check{Y}_2^N, \bar{V}_2^N | \underline{S}^N) \\
& = \mathbb{E}_{S^N} \left[ h_S(\check{Y}_2^N, \bar{V}_2^N) - h_S(\check{Y}_2^N, \bar{V}_2^N | W_2) \right] \\
& \stackrel{(a)}{=} \mathbb{E}_{S^N} \left[ h_S(Y_{22}^{M_2}, Y_{21}^{M_2}, \bar{Y}_2^{L_2}, \bar{V}_2^N) \right. \\
& \quad \left. - h_S(\check{V}_1^N, Z_1^N, Z_{22}^{M_2} | W_2) \right] \\
& = \mathbb{E}_{S^N} \left[ h_S(Y_{22}^{M_2}, \bar{Y}_2^{L_2} | Y_{21}^{M_2}, \bar{V}_2^N) \right] \\
& + \mathbb{E}_{S^N} \left[ h_S(Y_{21}^{M_2}, \bar{V}_2^N) - h_S(\check{V}_1^N, Z_1^N, Z_{22}^{M_2} | W_2) \right] \\
& \stackrel{(b)}{\leq} \mathbb{E}_{S^N} \left[ h_S(Y_{22}^{M_2} | \bar{V}_2^N) + h_S(\bar{Y}_2^{L_2} | \bar{V}_2^N) \right] \\
& + \mathbb{E}_{S^N} \left[ \sum_{t=1}^N h_S(\bar{V}_{2,t} | \bar{V}_2^{t-1}, Y_{21}^{(t-1)}) \right. \\
& + \sum_{t:S_{2,t}=1} h_S(Y_{21,t} | \bar{V}_2^{t-1}, Y_{21}^{(t-1)}) \\
& \quad \left. - \sum_{t=1}^N h_S(\check{V}_{1,t}, Z_{1,t}, Z_{22,t} | W_2, \check{V}_1^{t-1}, Z_1^{t-1}, Z_{22}^{(t-1)}) \right] \\
& \stackrel{(c)}{=} \mathbb{E}_{S^N} \left[ h_S(Y_{22}^{M_2} | \bar{V}_2^N) + h_S(\bar{Y}_2^{L_2} | \bar{V}_2^N) \right] \\
& + \mathbb{E}_{S^N} \left[ \sum_{t=1}^N h_S(\bar{V}_{2,t} | \bar{V}_2^{t-1}, Y_{21}^{(t-1)}) \right. \\
& + \sum_{t:S_{2,t}=1} h_S(Y_{21,t} | \bar{V}_2^{t-1}, Y_{21}^{(t-1)}) \Big] \\
& - \mathbb{E}_{S^N} \left[ \sum_{t=1}^N h_S(\check{V}_{1,t} | W_2, \check{V}_1^{t-1}, Z_1^{t-1}, Z_{22}^{(t-1)}) \right. \\
& + \sum_{t=1}^N h_S(Z_{1,t} | W_2, \check{V}_1^{t-1}, Z_1^{t-1}, Z_{22}^{(t-1)}) \\
& + \sum_{t:S_{2,t}=1} h_S(Z_{22,t} | W_2, \check{V}_1^{t-1}, Z_1^{t-1}, Z_{22}^{(t-1)}) \Big] \\
& \stackrel{(d)}{\leq} \mathbb{E}_{S^N} \left[ h_S(Y_{22}^{M_2} | \bar{V}_2^N) + h_S(\bar{Y}_2^{L_2} | \bar{V}_2^N) \right] \\
& + \mathbb{E}_{S^N} \left[ \sum_{t=1}^N h_S(\bar{V}_{2,t} | \bar{V}_2^{t-1}, Y_{21}^{(t-1)}) + \sum_{t:S_{2,t}=1} h_S(Y_{21,t}) \right]
\end{aligned}$$

$$\begin{aligned}
& - \mathbb{E}_{S^N} \left[ \sum_{t=1}^N h_S \left( \check{V}_{1,t} | W_2, \check{V}_1^{t-1}, Z_1^{t-1}, Z_{22}^{(t-1)} \right) \right] \\
& - N h(Z_1) - N p_2 h(Z_{22})
\end{aligned}$$

where (a) is by decomposing  $\check{Y}_2^N$  into  $(Y_{22}^{M_2}, Y_{21}^{M_2})$  for time slots where  $S_{2,t} = 1$ , and to  $\bar{Y}_2^{L_2}$  for time slots where  $S_{2,t} = 0$ , and by Lemma E.6. (b) is because conditioning reduces entropy and by chain rule. (c) is because for a given time slot  $t$ , the noise terms involved are independent from each other and from  $\check{V}_{1,t}$ . (d) is because conditioning reduces entropy and the noise processes are i.i.d., and because the noise terms are independent from the signals up to time  $t - 1$ .

Next, we consider the third mutual information term from (89).

$$\begin{aligned}
& I(W_1; \check{Y}_1^N, \check{V}_1^N | W_2, \underline{S}^N) \\
& = \mathbb{E}_{S^N} \left[ I_S(W_1; \check{Y}_1^N, \check{V}_1^N | W_2) \right] \\
& \stackrel{(a)}{\leq} \mathbb{E}_{S^N} \left[ I(W_1; \check{Y}_1^N, \check{V}_1^N | W_2, Z_{22}^{M_2}, Z_{11}^{M_1}) \right] \\
& = \mathbb{E}_{S^N} \left[ h_S(\check{Y}_1^N, \check{V}_1^N | W_2, Z_{22}^{M_2}, Z_{11}^{M_1}) \right] \\
& - \mathbb{E}_{S^N} \left[ h_S(\check{Y}_1^N, \check{V}_1^N | W_2, Z_{22}^{M_2}, Z_{11}^{M_1}, W_1) \right] \\
& = \sum_{t=1}^N \mathbb{E}_{S^N} \left[ h_S(\check{Y}_{1,t}, \check{V}_{1,t} | \check{Y}_1^{t-1}, \check{V}_1^{t-1}, W_2, Z_{22}^{M_2}, Z_{11}^{M_1}) \right] \\
& - \mathbb{E}_{S^N} \left[ h_S(\check{Y}_{1,t}, \check{V}_{1,t} | \check{Y}_1^{t-1}, \check{V}_1^{t-1}, W_2, Z_{22}^{M_2}, Z_{11}^{M_1}, W_1) \right] \\
& = \sum_{t=1}^N \mathbb{E}_{S^N} \left[ h_S(\check{Y}_{1,t} | \check{Y}_1^{t-1}, \check{V}_1^t, W_2, Z_{22}^{M_2}, Z_{11}^{M_1}) \right] \\
& + \mathbb{E}_{S^N} \left[ h_S(\check{V}_{1,t} | \check{Y}_1^{t-1}, \check{V}_1^{t-1}, W_2, Z_{22}^{M_2}, Z_{11}^{M_1}) \right] \\
& - \mathbb{E}_{S^N} \left[ h_S(\check{Y}_{1,t}, \check{V}_{1,t} | \check{Y}_1^{t-1}, \check{V}_1^{t-1}, W_2, Z_{22}^{M_2}, Z_{11}^{M_1}, W_1) \right] \\
& \stackrel{(b)}{=} \sum_{t=1}^N \mathbb{E}_{S^N} \left[ h_S(\check{Y}_{1,t} | \check{Y}_1^{t-1}, \check{V}_1^t, W_2, Z_{22}^{M_2}, Z_{11}^{M_1}, X_{2,t}) \right] \\
& + \mathbb{E}_{S^N} \left[ h_S(\check{V}_{1,t} | \check{Y}_1^{t-1}, \check{V}_1^{t-1}, W_2, Z_{22}^{M_2}, Z_{11}^{M_1}) \right] \\
& - \mathbb{E}_{S^N} \left[ h_S(\check{Y}_{1,t}, \check{V}_{1,t} \right. \\
& \quad \left. | \check{Y}_1^{t-1}, \check{V}_1^{t-1}, W_2, Z_{22}^{M_2}, Z_{11}^{M_1}, W_1, X_{1,t}, X_{2,t}) \right] \\
& \stackrel{(c)}{\leq} \sum_{t=1}^N \mathbb{E}_{S^N} \left[ h_S(\check{Y}_{1,t} | \check{V}_{1,t}, X_{2,t}) \right] \\
& + \mathbb{E}_{S^N} \left[ h_S(\check{V}_{1,t} | Y_{12}^{(t-1)}, \check{V}_1^{t-1}, W_2, Z_{22}^{M_2}, Z_{11}^{M_1}) \right] \\
& - \mathbb{E}_{S^N} \left[ \sum_{t: S_{2,t}=0} h_S(Z_{2,t}) + \sum_{t: S_{2,t}=1} h_S(Z_{21,t}) \right] \\
& - \mathbb{E}_{S^N} \left[ \sum_{t: S_{1,t}=0} h_S(Z_{1,t}) + \sum_{t: S_{1,t}=1} h_S(Z_{11,t}, Z_{12,t}) \right] \\
& \stackrel{(d)}{\leq} \sum_{t=1}^N \mathbb{E}_{S^N} \left[ h_S(\check{Y}_{1,t} | \check{V}_{1,t}, X_{2,t}) \right] \\
& + \mathbb{E}_{S^N} \left[ h_S(\check{V}_{1,t} | Y_{12}^{(t-1)}, \check{V}_1^{t-1}, W_2, Z_{22}^{(t-1)}, Z_{11}^{(t-1)}) \right]
\end{aligned}$$

$$\begin{aligned}
& - N(1 - p_1)h(Z_1) - N p_1 h(Z_{11}) - N p_1 h(Z_{12}) \\
& - N p_2 h(Z_{21}) - N(1 - p_2)h(Z_2)
\end{aligned}$$

where (a) follows by the fact that  $(Z_{11}^{M_1}, Z_{22}^{M_2})$  is independent from  $W_1$  given  $W_2$ , (b) is because  $X_{1,t} \stackrel{f}{=} (W_1, \check{Y}_1^{t-1}, \underline{S}^{t-1})$  and  $X_{2,t} \stackrel{f}{=} (W_2, \check{V}_1^{t-1}, Z_{22}^{(t-1)}, \underline{S}^{t-1})$ . In (c), the first two terms are upper bounded using the fact that  $\check{Y}_1^{t-1} = (Y_{11}^{(t-1)}, Y_{12}^{(t-1)}, \bar{Y}_1^{[t-1]})$ , and that conditioning reduces entropy. The noise terms are obtained by subtracting  $X_{1,t}$  and  $X_{2,t}$  from  $\check{Y}_{1,t}$  and  $\check{V}_{1,t}$ , and using the fact that noise variables at time  $t$  are independent from the variables up to time  $t - 1$ . (d) is because conditioning reduces entropy and because noise processes are i.i.d.

Putting everything back together, we have

$$\begin{aligned}
& N(2R_1 + R_2 - \epsilon_N) \\
& \leq \sum_{t=1}^N h(\bar{Y}_{1,t}) + \mathbb{E}_{S^N} \left[ h_S(Y_{22}^{M_2} | \bar{V}_2^N) + h_S(\bar{Y}_2^{L_2} | \bar{V}_2^N) \right] \\
& + \mathbb{E}_{S^N} \left[ \sum_{t=1}^N I_S(\bar{V}_{2,t}; W_1 | \bar{V}_2^{t-1}, Y_{21}^{(t-1)}) \right. \\
& + \sum_{t=1}^N I_S(\check{V}_{1,t}; Z_1^{t-1} | Y_{12}^{(t-1)}, \check{V}_1^{t-1}, W_2, Z_{22}^{(t-1)}, Z_{11}^{(t-1)}) \\
& + \sum_{t=1}^N h_S(\check{Y}_{1,t} | \check{V}_{1,t}, X_{2,t}) + \sum_{t: S_{2,t}=1} h_S(Y_{21,t}) \left. \right] \\
& - N(1 - p_1)h(Z_1) - N p_1 h(Z_{11}) - N p_1 h(Z_{12}) \\
& - N p_2 h(Z_{21}) - N(1 - p_2)h(Z_2) - N h(Z_1) \\
& - N p_2 h(Z_{22})
\end{aligned}$$

Using Lemma E.3, E.4, E.5, E.7, E.8 and E.9 to bound each of these terms, we get the desired bound.

### E. Lemmas

In this subsection, we prove the lemmas that have been used in the proofs of the previous subsections.

**Lemma E.1.**  $X_{2,t} \stackrel{f}{=} (W_2, \tilde{V}_1^{t-1}, \underline{S}^t)$

*Proof.* Note that

$$X_{2,1} \stackrel{f}{=} W_2$$

and by the definition of the channel,

$$\begin{aligned}
X_{2,t} & \stackrel{f}{=} (W_2, \tilde{Y}_2^{t-1}, \underline{S}^t) \\
& \stackrel{(a)}{\stackrel{f}{=}} (W_2, \tilde{V}_1^{t-1}, X_2^{t-1}, \underline{S}^t),
\end{aligned}$$

hence the result follows by induction on  $t$ . (a) follows because  $\tilde{Y}_2^{t-1} = S_2^{t-1} h_{22} X_2^{t-1} + \tilde{V}_1^{t-1}$ .  $\square$

**Lemma E.2.**

$$\sum_{t=1}^N h(Y_{1,t} | \tilde{V}_{1,t}, S_{2,t}, X_{2,t}) \leq N p_2 \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right)$$

$$+ N(1 - p_2) \log(1 + \text{SNR}_1)$$

*Proof.*

$$\begin{aligned} \sum_{t=1}^N h(Y_{1,t} | \tilde{V}_{1,t}, S_{2,t}, X_{2,t}) &= \sum_{t=1}^N p_2 h(Y_{1,t} | V_{1,t}, X_{2,t}) \\ &\quad + (1 - p_2) h(Y_{1,t} | X_{2,t}) \\ &\stackrel{(a)}{=} \sum_{t=1}^N p_2 h(Y_{1,Q} | V_{1,Q}, X_{2,Q}, Q = t) \\ &\quad + (1 - p_2) h(Y_{1,Q} | X_{2,Q}, Q = t) \\ &= N p_2 h(Y_{1,Q} | V_{1,Q}, X_{2,Q}, Q) \\ &\quad + N(1 - p_2) h(Y_{1,Q} | X_{2,Q}, Q) \\ &\stackrel{(b)}{=} N p_2 h(Y_1 | V_1, X_2, Q) + N(1 - p_2) h(Y_1 | X_2, Q) \\ &\stackrel{(c)}{\leq} N p_2 \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\ &\quad + N(1 - p_2) \log(1 + \text{SNR}_1) \end{aligned}$$

where (a) follows by introducing a time-sharing variable  $Q$  uniformly distributed between 1 and  $N$ , (b) is by defining  $Y_i := Y_{i,Q}$ ,  $X_i := X_{i,Q}$  and  $V_i := V_{i,Q}$ . (c) follows by the fact that choosing jointly Gaussian input distribution with correlation coefficient  $\rho = 0$  for  $p(x_1, x_2)$  maximizes the given conditional differential entropy.  $\square$

**Lemma E.3.**

$$\mathbb{E}_{S^N} \left[ \sum_{t:S_{i,t}=1} h_S(Y_{ij,t}) \right] = N p_i \log \left( 2\pi e \left( \frac{1}{2} + \text{INR}_i \right) \right)$$

for  $(i, j) = (1, 2), (2, 1)$ .

*Proof.*

$$\begin{aligned} \mathbb{E}_{S^N} \left[ \sum_{t:S_{i,t}=1} h_S(Y_{ij,t}) \right] &= \mathbb{E}_{S^N} \left[ M_i \frac{1}{M_i} \sum_{t:S_{i,t}=1} h_S(Y_{ij,t}) \right] \\ &\leq \mathbb{E}_{S^N} \left[ M_i \frac{1}{M_i} \sum_{t:S_{i,t}=1} \log 2\pi e \left( \frac{1}{2} + |h_{ij}|^2 P_{j,t} \right) \right] \\ &= \mathbb{E} \left[ M_i \mathbb{E} \left[ \frac{1}{M_i} \sum_{t:S_{i,t}=1} \log 2\pi e \left( \frac{1}{2} + |h_{ij}|^2 P_{j,t} \right) \middle| M_i \right] \right] \\ &\stackrel{(a)}{\leq} \mathbb{E}_{M_i} \left[ M_i \mathbb{E} \left[ \log 2\pi e \left( \frac{1}{2} + |h_{ij}|^2 P_j^{(i1)} \right) \middle| M_i \right] \right] \\ &\stackrel{(b)}{\leq} \mathbb{E}_{M_i} \left[ M_i \log 2\pi e \left( \frac{1}{2} + |h_{ij}|^2 \mathbb{E} [P_j^{(i1)} | M_i] \right) \right] \\ &\stackrel{(c)}{\leq} \mathbb{E}_{M_i} \left[ M_i \log 2\pi e \left( \frac{1}{2} + |h_{ij}|^2 P_j \right) \right] \\ &= \mathbb{E} [M_i] \log \left( 2\pi e \left( \frac{1}{2} + \text{INR}_i \right) \right) \\ &= N p_i \log \left( 2\pi e \left( \frac{1}{2} + \text{INR}_i \right) \right) \end{aligned}$$

where (a) and (b) follow by Jensen's inequality (since  $\log(\cdot)$  is concave), and (c) follows since  $P_j^{(i1)}$  averaged over the

realizations of  $S_i$  is the average power, which is less than the power constraint  $P_j$ .  $\square$

**Lemma E.4.**

$$\begin{aligned} \mathbb{E}_{S^N} \left[ I_S(V_{i,t}; W_j, Z_{jj}^{(t-1)}, Z_i^{[t-1]}, Z_{ij}^{(t-1)} | V_i^{t-1}, Y_{ij}^{(t-1)}) \right] \\ = \mathbb{E}_{S^N} \left[ I_S(V_{i,t}; Z_i^{t-1} | Y_{ij}^{(t-1)}, V_i^{t-1}, W_j, Z_{jj}^{(t-1)}, Z_{ii}^{(t-1)}) \right] \\ = 0 \end{aligned}$$

for  $(i, j) = (1, 2), (2, 1)$ .

*Proof.* Since all variables involved are related to  $V_{i,t}$  through  $X_{i,t}$ ; by data processing inequality,

$$\begin{aligned} \mathbb{E}_{S^N} \left[ I_S(V_{i,t}; W_j, Z_{jj}^{(t-1)}, Z_i^{[t-1]}, Z_{ij}^{(t-1)} | V_i^{t-1}, Y_{ij}^{(t-1)}) \right] \\ \leq \mathbb{E}_{S^N} \left[ I_S(X_{i,t}; W_j, Z_{jj}^{(t-1)}, Z_i^{[t-1]}, Z_{ij}^{(t-1)} | V_i^{t-1}, Y_{ij}^{(t-1)}) \right] \\ \leq \mathbb{E}_{S^N} \left[ I_S(W_i, Z_{ii}^{(t-1)}; W_j, Z_{jj}^{(t-1)}, Z_i^{[t-1]}, Z_{ij}^{(t-1)} \right. \\ \left. | V_i^{t-1}, Y_{ij}^{(t-1)}) \right] \end{aligned}$$

where the latter inequality follows by the fact that  $X_{i,t} \stackrel{f}{=} (W_i, Y_{ij}^{(t-1)}, Z_{ii}^{(t-1)}, \underline{S}^{t-1})$ . Similarly, the second mutual information can be bounded by

$$\begin{aligned} \mathbb{E}_{S^N} \left[ I_S(V_{i,t}; Z_i^{t-1} | Y_{ij}^{(t-1)}, V_i^{t-1}, W_j, Z_{jj}^{(t-1)}, Z_{ii}^{(t-1)}) \right] \\ \leq \mathbb{E}_{S^N} \left[ I_S(X_{i,t}; Z_i^{[t-1]}, Z_{ij}^{(t-1)} \right. \\ \left. | Y_{ij}^{(t-1)}, V_i^{t-1}, W_j, Z_{jj}^{(t-1)}, Z_{ii}^{(t-1)}) \right] \\ \leq \mathbb{E}_{S^N} \left[ I_S(W_i, Z_{ii}^{(t-1)}; Z_{jj}^{(t-1)}, Z_i^{[t-1]}, Z_{ij}^{(t-1)}, W_j \right. \\ \left. | V_i^{t-1}, Y_{ij}^{(t-1)}) \right] \end{aligned}$$

where the first step is because  $Z_i^{t-1} \stackrel{f}{=} (Z_i^{[t-1]}, Z_{ij}^{(t-1)}, Z_{ii}^{(t-1)}, \underline{S}^{t-1})$ , and second step is because for random variables  $A, B, C$ ;  $I(A, B; C) \geq I(A; C|B)$ ; and  $X_{i,t} \stackrel{f}{=} (W_i, Y_{ij}^{(t-1)}, Z_{ii}^{(t-1)}, \underline{S}^{t-1})$ . Note that we have the same upper bound for both mutual information terms. We will next show that this upper bound is zero.

To show conditional independence, we will use the property that  $X$  and  $Y$  are independent given  $Z$  if and only if the probability distribution  $p(X, Y, Z)$  can be factorized as

$$p(X, Y, Z) = f(X, Z)g(Y, Z)$$

for some functions  $f$  and  $g$ . Consider the joint distribution of all the variables involved in the above mutual information (we define  $p_S(\cdot) := p(\cdot | \underline{S}^N = S^N)$ ).

$$\begin{aligned} p_S(W_i, Z_{ii}^{(t-1)}, Z_{jj}^{(t-1)}, Z_i^{[t-1]}, Z_{ij}^{(t-1)}, W_j, V_i^{t-1}, Y_{ij}^{(t-1)}) &= \\ p(W_i) p(W_j) \prod_{\tau=1}^{t-1} p_S(Z_{ii,(\tau)}, Z_{jj,(\tau)}, Z_{i,(\tau)}, Z_{ij,(\tau)}, V_{i,\tau}, Y_{ij,(\tau)}) \\ &\quad | Z_{ii}^{(\tau-1)}, Z_{jj}^{(\tau-1)}, Z_i^{[\tau-1]}, Z_{ij}^{(\tau-1)}, V_i^{\tau-1}, Y_{ij}^{(\tau-1)}, W_i, W_j \\ &\stackrel{(a)}{=} p(W_i) p(W_j) \prod_{\tau=1}^{t-1} p_S(Z_{ii,(\tau)}) p_S(Z_{jj,(\tau)}) p_S(Z_{i,(\tau)}) \\ &\quad \cdot p_S(Z_{ij,(\tau)}) p_S(V_{i,\tau} | Z_{ii}^{(\tau-1)}, Y_{ij}^{(\tau-1)}, W_i) \end{aligned}$$



$$\begin{aligned}
& \cdot p_S(Y_{ij,(\tau)}|Z_{jj}^{(\tau-1)}, Z_{ij}^{(\tau)}, V_i^{\tau-1}, W_j) \\
& = f(W_i, Z_{ii}^{(t-1)}, V_i^{t-1}, Y_{ij}^{(t-1)}) \\
& \cdot g(Z_{jj}^{(t-1)}, Z_i^{(t-1)}, Z_{ij}^{(t-1)}, W_j, V_i^{t-1}, Y_{ij}^{(t-1)})
\end{aligned}$$

where (a) follows since

$$\begin{aligned}
Y_{ij,(\tau)} & \stackrel{f}{=} (X_{j,\tau}, Z_{ij}^{(\tau)}, \underline{S}^\tau) \\
& \stackrel{f}{=} (Z_{jj}^{(\tau-1)}, Z_{ij}^{(\tau)}, V_i^{\tau-1}, \underline{S}^\tau, W_j)
\end{aligned}$$

and

$$\begin{aligned}
V_{i,\tau} & \stackrel{f}{=} (X_{i,\tau}, Z_{j,(\tau)}, Z_{ji,(\tau)}) \\
& \stackrel{f}{=} (Z_{ii}^{(\tau-1)}, Y_{ij}^{(\tau-1)}, \underline{S}^{\tau-1}, W_i, Z_{j,(\tau)}, Z_{ji,(\tau)})
\end{aligned}$$

and  $(Z_{j,(\tau)}, Z_{ji,(\tau)})$  is independent of everything else. In the last line, we define

$$\begin{aligned}
& f(W_i, Z_{ii}^{(t-1)}, V_i^{t-1}, Y_{ij}^{(t-1)}) \\
& = p(W_i) \prod_{\tau=1}^{t-1} p_S(Z_{ii,(\tau)}) p_S(V_{i,\tau}|Z_{ii}^{(\tau-1)}, Y_{ij}^{(\tau-1)}, W_i) \\
& g(Z_{jj}^{(t-1)}, Z_i^{(t-1)}, Z_{ij}^{(t-1)}, W_j, V_i^{t-1}, Y_{ij}^{(t-1)}) \\
& = p(W_j) \prod_{\tau=1}^{t-1} p_S(Z_{jj,(\tau)}) p_S(Z_{i,(\tau)}) p_S(Z_{ij,(\tau)}) \\
& \cdot p_S(Y_{ij,\tau}|Z_{jj}^{(\tau-1)}, Z_{ij}^{(\tau)}, V_i^{\tau-1}, W_j)
\end{aligned}$$

from which the result follows.

**Lemma E.5.**

$$\mathbb{E}_{S^N} [I_S(\bar{V}_{j,t}; W_i | \bar{V}_j^{t-1}, Y_{ji}^{(t-1)})] = 0$$

*Proof.*

$$\begin{aligned}
& \mathbb{E}_{S^N} [I_S(\bar{V}_{j,t}; W_i | \bar{V}_j^{t-1}, Y_{ji}^{(t-1)})] \\
& \leq \mathbb{E}_{S^N} [I_S(X_{j,t}; W_i | \bar{V}_j^{t-1}, Y_{ji}^{(t-1)})] \\
& \leq \mathbb{E}_{S^N} [I_S(W_j, Z_{jj}^{(t-1)}; W_i | \bar{V}_j^{t-1}, Y_{ji}^{(t-1)})]
\end{aligned}$$

where the first step follows by data processing inequality, and the second one follows by  $X_{j,t} \stackrel{f}{=} (W_j, Y_{ji}^{(t-1)}, Z_{jj}^{(t-1)}, \underline{S}^{t-1})$ . The proof technique is similar to that of Lemma E.4. The probability distribution of the involved variables is

$$\begin{aligned}
& p(W_i, W_j, Z_{jj}^{(t-1)}, \bar{V}_j^{t-1}, Y_{ji}^{(t-1)}) \\
& = p(W_i) p(W_j) \prod_{\tau=1}^{t-1} p_S(Z_{jj,(\tau)}) \\
& \quad p_S(\bar{V}_{j,\tau}, Y_{ji,(\tau)} | \bar{V}_j^{\tau-1}, Y_{ji}^{(\tau-1)}, Z_{jj}^{(\tau-1)}, W_i, W_j) \\
& \stackrel{(a)}{=} p(W_i) p(W_j) \prod_{\tau=1}^{t-1} p_S(Z_{jj,(\tau)}) \\
& \quad p_S(\bar{V}_{j,\tau} | Y_{ji}^{(\tau-1)}, Z_{jj}^{(\tau-1)}, W_j) p_S(Y_{ji,(\tau)} | W_i, \bar{V}_j^{t-1})
\end{aligned}$$

where (a) follows by the fact that

$$\bar{V}_{j,\tau} \stackrel{f}{=} (W_j, Y_{ji}^{(\tau-1)}, Z_{jj}^{(t-1)}, \underline{S}^{t-1}, Z_{i,\tau}),$$

$$Y_{ji,(\tau)} \stackrel{f}{=} (W_i, \bar{V}_j^{t-1}, \underline{S}^{t-1}, Z_{ji,(\tau)})$$

and that  $Z_{i,\tau}$  and  $Z_{ji,(\tau)}$  are independent of everything else. Then the result follows by defining

$$\begin{aligned}
& f(W_i, \bar{V}_j^{t-1}, Y_{ji}^{(t-1)}) \\
& = p(W_i) \prod_{\tau=1}^{t-1} p_S(Y_{ji,(\tau)} | W_i, \bar{V}_j^{t-1}) \\
& g(W_j, Z_{jj}^{(t-1)}, \bar{V}_j^{t-1}, Y_{ji}^{(t-1)}) \\
& = p(W_j) \prod_{\tau=1}^{t-1} p_S(Z_{jj,(\tau)}) p_S(\bar{V}_{j,\tau} | Y_{ji}^{(\tau-1)}, Z_{jj}^{(\tau-1)}, W_i)
\end{aligned}$$

and noting that the above probability distribution factorizes as  $f \cdot g$ .  $\square$

**Lemma E.6.**

$$\begin{aligned}
& \mathbb{E}_{S^N} [h_S(\check{Y}_i^N, \check{V}_i^N | W_i)] \\
& = \mathbb{E}_{S^N} [h_S(\check{V}_j^N, Z_{ii}^{M_i}, Z_j^{L_j}, Z_{ji}^{M_j} | W_i)] \\
& \mathbb{E}_{S^N} [h_S(\check{Y}_i^N, \bar{V}_i^N | W_i)] \\
& = \mathbb{E}_{S^N} [h_S(\check{V}_j^N, Z_{ii}^{M_i}, Z_j^N | W_i)]
\end{aligned}$$

for  $(i, j) = (1, 2), (2, 1)$ .

*Proof.*

$$\begin{aligned}
& \mathbb{E}_{S^N} [h_S(\check{Y}_i^N, \check{V}_i^N | W_i)] \\
& = \mathbb{E}_{S^N} \left[ \sum_{t=1}^N h_S(\check{Y}_{i,t}, \check{V}_{i,t} | W_i, \check{Y}_i^{t-1}, \check{V}_i^{t-1}) \right] \\
& \stackrel{(a)}{=} \mathbb{E}_{S^N} \left[ \sum_{t=1}^N h_S(\check{Y}_{i,t}, \check{V}_{i,t} | W_i, \check{Y}_i^{t-1}, \check{V}_i^{t-1}, X_i^t) \right] \\
& = \mathbb{E}_{S^N} \left[ \sum_{t=1}^N h_S(\bar{Y}_{i,[t]}, Y_{ii,(t)}, Y_{ij,(t)}, \bar{V}_{i,[t]}, Y_{ji,(t)} \right. \\
& \quad \left. | W_i, \bar{Y}_i^{[t-1]}, Y_{ii}^{(t-1)}, Y_{ij}^{(t-1)}, \bar{V}_i^{[t-1]}, Y_{ji}^{(t-1)}, X_i^t) \right] \\
& = \mathbb{E}_{S^N} \left[ \sum_{t=1}^N h_S(\bar{V}_{j,[t]}, Z_{ii,(t)}, Y_{ij,(t)}, Z_{j,[t]}, Z_{ji,(t)} \right. \\
& \quad \left. | W_i, \bar{V}_j^{[t-1]}, Z_{ii}^{(t-1)}, Y_{ij}^{(t-1)}, Z_j^{[t-1]}, Z_{ji}^{(t-1)}) \right] \\
& = \mathbb{E}_{S^N} [h_S(\bar{V}_j^{L_i}, Z_{ii}^{M_i}, Y_{ij}^{M_i}, Z_j^{L_j}, Z_{ji}^{M_j} | W_i)] \\
& \stackrel{(b)}{=} \mathbb{E}_{S^N} [h_S(\check{V}_j^N, Z_{ii}^{M_i}, Z_j^{L_j}, Z_{ji}^{M_j} | W_i)]
\end{aligned}$$

where (a) is because  $X_i^t \stackrel{f}{=} (W_i, \check{Y}_i^{t-1}, \check{S}^{t-1})$ , and (b) is because  $\check{V}_j^N = (\bar{V}_j^{L_i}, Y_{ij}^{M_i})$ . The second equality can be proved using similar steps.  $\square$

**Lemma E.7.**

$$\begin{aligned}
& \mathbb{E}_{S^N} [h_S(\bar{Y}_i^{L_i} | \check{V}_i^N)] \leq a \\
& \mathbb{E}_{S^N} [h_S(\bar{Y}_i^{L_i} | \bar{V}_i^N)] \leq a
\end{aligned}$$

for  $(i, j) = (1, 2), (2, 1)$ , where

$$a = N(1 - p_i) \cdot \log 2\pi e \left( 1 + \text{INR}_i + \frac{\text{SNR}_i + 2\sqrt{\text{SNR}_i \cdot \text{INR}_i}}{1 + \text{INR}_j} \right)$$

*Proof.*

$$\begin{aligned} & \mathbb{E}_{S^N} \left[ h_S \left( \bar{Y}_i^{L_i} | \check{V}_i^N \right) \right] \\ & \leq \mathbb{E}_{S^N} \left[ \sum_{t: S_{i,t}=0} h_S \left( \bar{Y}_{i,t} | \check{V}_{i,t} \right) \right] \\ & \stackrel{(a)}{=} \mathbb{E}_{S^N} \left[ L_i \frac{1}{L_i} \sum_{t: S_{i,t}=0} h_S \left( \bar{Y}_{i,t} | \check{V}_{i,t}, Q = t \right) \right] \\ & \stackrel{(b)}{=} \mathbb{E}_{S^N} \left[ L_i h_S \left( \bar{Y}_i | \check{V}_i, Q \right) \right] \\ & = \mathbb{E}_{S^N} \left[ L_i \left( p_j h_S \left( \bar{Y}_i | Y_{ji}, Q \right) + (1 - p_j) h_S \left( \bar{Y}_i | \bar{V}_i, Q \right) \right) \right] \\ & \stackrel{(c)}{\leq} \mathbb{E} \left[ L_i p_j \log 2\pi e \left( 1 + \text{INR}_i + \frac{\text{SNR}_i + 2\sqrt{\text{SNR}_i \cdot \text{INR}_i}}{1 + 2\text{INR}_j} \right) \right. \\ & \quad \left. + L_i (1 - p_j) \right. \\ & \quad \left. \cdot \log 2\pi e \left( 1 + \text{INR}_i + \frac{\text{SNR}_i + 2\sqrt{\text{SNR}_i \cdot \text{INR}_i}}{1 + \text{INR}_j} \right) \right] \\ & \leq (1 - p_i) \log 2\pi e \left( 1 + \text{INR}_i + \frac{\text{SNR}_i + 2\sqrt{\text{SNR}_i \cdot \text{INR}_i}}{1 + \text{INR}_j} \right) \end{aligned}$$

where (a) is by introducing a time-sharing random variable  $Q$  with uniform distribution over the set  $\{t : S_{i,t} = 0\}$ , and (b) follows by setting  $\bar{Y}_i = \bar{Y}_{i,Q}$  and  $\check{V}_i = \check{V}_{i,Q}$ . (c) follows by the fact that choosing jointly Gaussian input distribution with correlation coefficient  $\rho = 0$  for  $p(x_1, x_2)$  maximizes the given conditional differential entropy.

By following similar steps, we can show that

$$\begin{aligned} & \mathbb{E}_{S^N} \left[ h_S \left( \bar{Y}_i^{L_i} | \bar{V}_i^N \right) \right] \leq \mathbb{E}_{S^N} \left[ L_i h_S \left( \bar{Y}_i | \bar{V}_i, Q \right) \right] \\ & \leq (1 - p_i) \log 2\pi e \left( 1 + \text{INR}_i + \frac{\text{SNR}_i + 2\sqrt{\text{SNR}_i \cdot \text{INR}_i}}{1 + \text{INR}_j} \right) \end{aligned}$$

□

**Lemma E.8.**

$$\begin{aligned} & \mathbb{E}_{S^N} \left[ h_S \left( Y_{ii}^{M_i} | \check{V}_i^N \right) \right] \leq b \\ & \mathbb{E}_{S^N} \left[ h_S \left( Y_{ii}^{M_i} | \bar{V}_i^N \right) \right] \leq b \end{aligned}$$

where

$$b = N p_i \log 2\pi e \left( \frac{1}{2} + \frac{\text{SNR}_i}{2\text{INR}_j + 1} \right)$$

for  $(i, j) = (1, 2), (2, 1)$ .

*Proof.*

$$\begin{aligned} & \mathbb{E}_{S^N} \left[ h_S \left( Y_{ii}^{M_i} | \check{V}_i^N \right) \right] \leq \mathbb{E}_{S^N} \left[ \sum_{t: S_{i,t}=1} h_S \left( Y_{ii,t} | \check{V}_{i,t} \right) \right] \\ & \stackrel{(a)}{=} \mathbb{E}_{S^N} \left[ M_i \frac{1}{M_i} \sum_{t: S_{i,t}=1} h_S \left( Y_{ii,t} | \check{V}_{i,t}, Q = t \right) \right] \end{aligned}$$

$$\begin{aligned} & \stackrel{(b)}{=} \mathbb{E}_{S^N} \left[ M_i h_S \left( Y_{ii} | \check{V}_i, Q \right) \right] \\ & = \mathbb{E}_{S^N} \left[ M_i \left( (1 - p_j) h_S \left( Y_{ii} | \check{V}_i, Q \right) + p_j h_S \left( Y_{ii} | Y_{ji}, Q \right) \right) \right] \\ & \stackrel{(c)}{\leq} \mathbb{E}_{S^N} \left[ M_i \left( (1 - p_j) \log 2\pi e \left( \frac{1}{2} + \frac{2\text{SNR}_i + \frac{1}{2}}{2\text{INR}_j + 1} \right) \right. \right. \\ & \quad \left. \left. + p_j \log 2\pi e \left( \frac{1}{2} + \frac{\text{SNR}_i}{2\text{INR}_j + 1} \right) \right) \right] \\ & \leq p_i \log 2\pi e \left( \frac{1}{2} + \frac{\text{SNR}_i}{2\text{INR}_j + 1} \right) \end{aligned}$$

where (a) is by introducing a time-sharing random variable  $Q$  with uniform distribution over the set  $\{t : S_{i,t} = 1\}$ , and (b) follows by setting  $Y_{ii} = Y_{ii,Q}$  and  $\check{V}_i = \check{V}_{i,Q}$ . (c) follows by the fact that choosing jointly Gaussian input distribution with correlation coefficient  $\rho = 0$  for  $p(x_1, x_2)$  maximizes the given conditional differential entropy. Similarly,

$$\begin{aligned} & \mathbb{E}_{S^N} \left[ h_S \left( Y_{ii}^{M_i} | \bar{V}_i^N \right) \right] \leq \mathbb{E}_{S^N} \left[ M_i h_S \left( Y_{ii} | \bar{V}_i, Q \right) \right] \\ & \leq p_i \log 2\pi e \left( \frac{1}{2} + \frac{\text{SNR}_i}{2\text{INR}_j + 1} \right) \end{aligned}$$

□

**Lemma E.9.**

$$\begin{aligned} & \mathbb{E}_{S^N} \left[ \sum_{t=1}^N h_S \left( \check{Y}_{i,t} | \check{V}_{i,t}, X_{j,t} \right) \right] \\ & \leq p_i \log 2\pi e \left( \frac{1}{2} + \frac{\text{SNR}_i}{2\text{INR}_j + 1} \right) + p_i \log 2\pi e \frac{1}{2} \\ & \quad + (1 - p_i) \log 2\pi e \left( 1 + \frac{\text{SNR}_i}{1 + \text{INR}_j} \right) \end{aligned}$$

for  $(i, j) = (1, 2), (2, 1)$ .

*Proof.*

$$\begin{aligned} & \mathbb{E}_{S^N} \left[ \sum_{t=1}^N h_S \left( \check{Y}_{i,t} | \check{V}_{i,t}, X_{j,t} \right) \right] \\ & \stackrel{(a)}{=} \mathbb{E}_{S^N} \left[ N \frac{1}{N} \sum_{t=1}^N h_S \left( \check{Y}_{i,t} | \check{V}_{i,t}, X_{j,t}, Q = t \right) \right] \\ & = \mathbb{E}_{S^N} \left[ N h_S \left( \check{Y}_{i,Q} | \check{V}_{i,Q}, X_{j,Q}, Q \right) \right] \\ & \stackrel{(b)}{=} N h(\check{Y}_i | \check{V}_i, X_j, Q) \\ & = p_i h(Y_{ii}, Y_{ij} | \check{V}_i, X_j, Q) + (1 - p_i) h(\bar{Y}_i | \check{V}_i, X_j, Q) \\ & \leq p_i h(Y_{ii} | \check{V}_i) + p_i h(Y_{ij} | X_j) \\ & \quad + (1 - p_i) \left[ (1 - p_j) h(\bar{Y}_i | \check{V}_i, X_j) + p_j h(\bar{Y}_i | Y_{ji}, X_j) \right] \\ & \stackrel{(c)}{\leq} p_i \log 2\pi e \left( \frac{1}{2} + \frac{\text{SNR}_i}{2\text{INR}_j + 1} \right) + p_i \log 2\pi e \frac{1}{2} \\ & \quad + (1 - p_i) \log 2\pi e \left( 1 + \frac{\text{SNR}_i}{1 + \text{INR}_j} \right) \end{aligned}$$

where (a) is by introducing a uniformly distributed time-sharing random variable  $Q$ , and (b) is by defining  $\check{Y}_i = \check{Y}_{i,Q}$ ,  $\check{V}_i = \check{V}_{i,Q}$  and  $X_j = X_{j,Q}$ . (c) follows by the fact that choosing jointly Gaussian input distribution with correlation coefficient  $\rho = 0$  for  $p(x_1, x_2)$  maximizes the given conditional differential entropy. □

## APPENDIX F GAP ANALYSIS

In this section, we give upper bounds for the gap terms  $\delta_1$  and  $\delta_2$  from Theorem 3.2. We also compare our achievable region with the outer bound of [1] for the case  $p_1 = p_2 = 1$ .

### A. Bounding $\delta_1$

We will show that, each of the bounds (79), (80), and (81) are within a constant gap of the region given in (11)–(14). Without loss of generality, we focus on the case  $(i, j) = (1, 2)$ , and start with the first bound in (79).

$$\begin{aligned}
 A_1 + B_2 &= \log \left( 3 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
 &\quad + \log (2 + \text{INR}_2) - 2 \log 3 - C_1 - C_2 \\
 &\geq \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
 &\quad + \log (1 + \text{INR}_2) - 2 \log 3 - C_1 - C_2 \\
 &= \log (1 + \text{SNR}_1 + \text{INR}_2) - 2 \log 3 - C_1 - C_2 \\
 &= \log (1 + \text{SNR}_1) + \log \left( 1 + \frac{\text{INR}_2}{1 + \text{SNR}_1} \right) \\
 &\quad - 2 \log 3 - C_1 - C_2 \\
 &\geq (11)_{(1,2),R} - 2 \log 3 - C_1 - C_2
 \end{aligned}$$

where  $(11)_{(1,2),R}$  refers to bound on the right-hand side of (11), evaluated with  $(i, j) = (1, 2)$ . Next, we consider the second bound in (79). If  $\text{SNR}_1 \geq \text{INR}_1$ ,

$$\begin{aligned}
 D_1 &= \log (3 + \text{SNR}_1) - \log 3 - C_1 \\
 &\geq \log (1 + \text{SNR}_1) - \log 3 - C_1 \\
 &\geq \log (1 + \text{SNR}_1 + \text{INR}_1) - \log 3 - C_1 - 1 \\
 &= (11)_{(1,2),L} - \log 3 - C_1 - 1
 \end{aligned}$$

where  $(11)_{(1,2),L}$  refers to bound on the left-hand side of (11), evaluated with  $(i, j) = (1, 2)$ . If  $\text{SNR}_1 < \text{INR}_1$ ,

$$\begin{aligned}
 D_1 &= \log (3 + \text{SNR}_1) + p_2 \log \left( 1 + \frac{\text{INR}_2}{3 + \text{SNR}_1} \right) \\
 &\quad - p_2 \log \frac{5}{3} - \log 3 - C_1 \\
 &\geq \log (1 + \text{SNR}_1) + p_2 \log \left( 1 + \frac{\text{INR}_2}{1 + \text{SNR}_1} \right) \\
 &\quad - p_2 \log \frac{5}{3} - p_2 \log 3 - \log 3 - C_1 \\
 &= (11)_{(1,2),R} - p_2 \log 5 - \log 3 - C_1
 \end{aligned}$$

Next, we consider the first bound in (80).

$$\begin{aligned}
 A_1 + G_2 &= \log \left( 3 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
 &\quad + \log (2 + \text{SNR}_2 + \text{INR}_2) - 2 \log 3 - C_1 - C_2 - \kappa_1 \\
 &\geq \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
 &\quad + \log (1 + \text{SNR}_2 + \text{INR}_2) - 2 \log 3 - C_1 - C_2 - \kappa_1 \\
 &= (12)_{(1,2)} - 2 \log 3 - C_1 - C_2 - \kappa_1
 \end{aligned}$$

where  $(12)_{(1,2)}$  refers to the bound (12) evaluated with  $(i, j) = (1, 2)$ . For the bound  $F_1 + F_2$ , we first consider the case when  $\text{INR}_1 > \text{SNR}_1$ .

$$\begin{aligned}
 F_1 + F_2 &\geq \log \left( 2 + \text{INR}_1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
 &\quad + \log \left( 2 + \text{INR}_2 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right) - 2 \log 3 - C_1 - C_2 \\
 &\geq \log (1 + \text{INR}_1) + \log \left( 1 + \text{INR}_2 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right) \\
 &\quad - 2 \log 3 - C_1 - C_2 \\
 &\geq \log (1 + \text{SNR}_1 + \text{INR}_1) + \log \left( 1 + \text{INR}_2 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right) \\
 &\quad - 2 \log 3 - C_1 - C_2 - 1 \\
 &= (12)_{(2,1)} - 2 \log 3 - C_1 - C_2 - 1
 \end{aligned}$$

By symmetry, we can show that when  $\text{INR}_2 > \text{SNR}_2$ ,

$$F_1 + F_2 \geq (12)_{(1,2)} - 2 \log 3 - C_1 - C_2 - 1$$

Next we consider the only remaining case of  $\text{INR}_1 \leq \text{SNR}_1$ ,  $\text{INR}_2 \leq \text{SNR}_2$ .

$$\begin{aligned}
 F_1 + F_2 &= \log \left( 2 + \text{INR}_1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
 &\quad + \log \left( 2 + \text{INR}_2 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right) \\
 &\quad + p_1 \log \left( \frac{(2 + \text{INR}_1) \left( 3 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right)}{2 + \frac{\text{SNR}_1}{1 + \text{INR}_2} + \text{INR}_1} \right) \\
 &\quad + p_2 \log \left( \frac{(2 + \text{INR}_2) \left( 3 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right)}{2 + \frac{\text{SNR}_2}{1 + \text{INR}_1} + \text{INR}_2} \right) \\
 &\quad - 2 \log 3 - 2C_1 - 2C_2 - (p_1 + p_2) \log 6 \\
 &\stackrel{(a)}{\geq} \log \left( 1 + \text{INR}_1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
 &\quad + \log \left( 1 + \text{INR}_2 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right) \\
 &\quad + p_1 \log \left( \frac{(1 + \text{INR}_1) \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right)}{1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} + \text{INR}_1} \right) \\
 &\quad + p_2 \log \left( \frac{(1 + \text{INR}_2) \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right)}{1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} + \text{INR}_2} \right) \\
 &\quad - 2 \log 3 - 2C_1 - 2C_2 - (p_1 + p_2) \log 6 \\
 &= (13) - 2 \log 3 - 2C_1 - 2C_2 - (p_1 + p_2) \log 6
 \end{aligned}$$

where in (a), we used the fact that the function  $\log \left( \frac{x+a}{x+a+b} \right)$  is monotonically increasing in  $x$ , for  $x, a, b > 0$ . Finally, we consider the bound (81). Again, we distinguish the cases  $\text{INR}_2 > \text{SNR}_2$  and  $\text{INR}_2 \leq \text{SNR}_2$ . For the former case,

$$\begin{aligned}
 A_1 + F_2 + G_1 &= \log \left( 3 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
 &\quad + \log \left( 1 + \text{INR}_2 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right)
 \end{aligned}$$

$$\begin{aligned}
& + \log(2 + \text{SNR}_1 + \text{INR}_1) - 3 \log 3 - 2C_1 - C_2 - \kappa_2 \\
& \geq \log \left( 3 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
& + \log(1 + \text{INR}_2) + \log(2 + \text{SNR}_1 + \text{INR}_1) \\
& - 3 \log 3 - 2C_1 - C_2 - \kappa_2 \\
& \geq \log \left( 3 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
& + \log(1 + \text{INR}_2 + \text{SNR}_2) + \log(2 + \text{SNR}_1 + \text{INR}_1) \\
& - 3 \log 3 - 2C_1 - C_2 - \kappa_2 - 1 \\
& \geq \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
& + \log(1 + \text{INR}_2 + \text{SNR}_2) + \log(1 + \text{SNR}_1 + \text{INR}_1) \\
& - 3 \log 3 - 2C_1 - C_2 - \kappa_2 - 1 \\
& = (12)_{(1,2)} + (11)_{(1,2),L} - 3 \log 3 - 2C_1 - C_2 - \kappa_2 - 1
\end{aligned}$$

For the case  $\text{INR}_2 \leq \text{SNR}_2$ ,

$$\begin{aligned}
A_1 + F_2 + G_1 &= \log \left( 3 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
& + \log \left( 1 + \text{INR}_2 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right) + \log(2 + \text{SNR}_1 + \text{INR}_1) \\
& + p_2 \log \left( \frac{(2 + \text{INR}_2) \left( 3 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right)}{2 + \frac{\text{SNR}_2}{1 + \text{INR}_1} + \text{INR}_2} \right) \\
& - 3 \log 3 - 2C_1 - C_2 - p_2 \log 6 - \kappa_2 \\
& \stackrel{(a)}{\geq} \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
& + \log \left( 1 + \text{INR}_2 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right) + \log(1 + \text{SNR}_1 + \text{INR}_1) \\
& + p_2 \log \left( \frac{(1 + \text{INR}_2) \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right)}{1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} + \text{INR}_2} \right) \\
& - 3 \log 3 - 2C_1 - C_2 - p_2 \log 6 - \kappa_2 \\
& = (14) - 3 \log 3 - 2C_1 - 2C_2 - p_2 \log 6 - \kappa_2
\end{aligned}$$

where in (a), as before, we used the fact that the function  $\log \left( \frac{x+a}{x+a+b} \right)$  is monotonically increasing in  $x$ , for  $x, a, b > 0$ . By symmetry, similar gaps apply to the case  $(i, j) = (2, 1)$ . Now, we can upper bound  $\delta_1$  by noting that it cannot be larger than the maximum of the gaps found above (after proper normalization, *e.g.*, the gap found for the bound on  $R_1 + R_2$  is divided by 2, and the one on  $2R_1 + R_2$  is divided by 3). Hence, using the fact that  $C_i = 2p_j + p_i$  and  $\kappa_i = p_i$ , we find

$$\delta_1 < 2 \log 3 + 3(p_1 + p_2) \text{ bits.}$$

### B. Bounding $\delta_2$

In order to bound  $\delta_2$ , we compare the bounds obtained in Section VI with the bounds (11)–(14) one by one. Without loss of generality, we focus on  $(i, j) = (1, 2)$ , and begin with the first bound in (11).

$$\begin{aligned}
(11)_{(1,2),L} &= \log(1 + \text{SNR}_1 + \text{INR}_1) \\
&\geq \log \left( 1 + \text{SNR}_1 + \text{INR}_1 + 2\sqrt{\text{SNR}_1 \cdot \text{INR}_1} \right) - \log 3
\end{aligned}$$

$$\geq (51) - \log 3$$

Next, we consider the second bound in (11), and note that (53) = (11)<sub>R</sub>, where (11)<sub>R</sub> refers to the right-hand side of the minimization in (11). We now consider the bound (12).

$$\begin{aligned}
(12) &= \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
& + \log(1 + \text{SNR}_2 + \text{INR}_2) \\
&\geq \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
& + \log \left( 1 + \text{SNR}_2 + \text{INR}_2 + 2\sqrt{\text{SNR}_2 \cdot \text{INR}_2} \right) - \log 3 \\
&\geq (55) - \log 3
\end{aligned}$$

Let us take (13).

$$\begin{aligned}
(13) &= \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} + \text{INR}_1 \right) \\
& + \log \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} + \text{INR}_2 \right) \\
& + p_1 \log \left( \frac{(1 + \text{INR}_1) \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right)}{1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} + \text{INR}_1} \right) \\
& + p_2 \log \left( \frac{(1 + \text{INR}_2) \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right)}{1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} + \text{INR}_2} \right) \\
&\geq \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} + \text{INR}_1 + 2\sqrt{\text{SNR}_1 \cdot \text{INR}_1} \right) \\
& + \log \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} + \text{INR}_2 + 2\sqrt{\text{SNR}_2 \cdot \text{INR}_2} \right) \\
& + p_1 \log \left( \frac{(1 + \text{INR}_1) \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right)}{1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} + \text{INR}_1} \right) \\
& + p_2 \log \left( \frac{(1 + \text{INR}_2) \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right)}{1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} + \text{INR}_2} \right) \\
& - 2 \log 3 \\
&\geq \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} + \text{INR}_1 + 2\sqrt{\text{SNR}_1 \cdot \text{INR}_1} \right) \\
& + \log \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} + \text{INR}_2 + 2\sqrt{\text{SNR}_2 \cdot \text{INR}_2} \right) \\
& + p_1 \log \left( \frac{(1 + \text{INR}_1) \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right)}{1 + \frac{\text{SNR}_1 + 2\sqrt{\text{SNR}_1 \cdot \text{INR}_1}}{1 + \text{INR}_2} + \text{INR}_1} \right) \\
& + p_2 \log \left( \frac{(1 + \text{INR}_2) \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right)}{1 + \frac{\text{SNR}_2 + 2\sqrt{\text{SNR}_2 \cdot \text{INR}_2}}{1 + \text{INR}_1} + \text{INR}_2} \right) \\
& - 2 \log 3 \\
&\geq (59) - 2 \log 3 - 2p_1 - 2p_2
\end{aligned}$$

Finally, we consider (14)

$$\begin{aligned}
(14) &= \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
& + \log \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} + \text{INR}_2 \right)
\end{aligned}$$

$$\begin{aligned}
& + \log(1 + \text{SNR}_1 + \text{INR}_1) \\
& + p_2 \log \left( \frac{(1 + \text{INR}_2) \left(1 + \frac{\text{SNR}_2}{1 + \text{INR}_1}\right)}{1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} + \text{INR}_2} \right) \\
& \geq \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
& + \log \left( 1 + \frac{\text{SNR}_2 + 2\sqrt{\text{SNR}_2 \cdot \text{INR}_2}}{1 + \text{INR}_1} + \text{INR}_2 \right) \\
& + \log \left( 1 + \text{SNR}_1 + \text{INR}_1 + 2\sqrt{\text{SNR}_1 \cdot \text{INR}_1} \right) \\
& + p_2 \log \left( \frac{(1 + \text{INR}_2) \left(1 + \frac{\text{SNR}_2}{1 + \text{INR}_1}\right)}{1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} + \text{INR}_2} \right) \\
& - 2 \log 3 \\
& \geq \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
& + \log \left( 1 + \frac{\text{SNR}_2 + 2\sqrt{\text{SNR}_2 \cdot \text{INR}_2}}{1 + \text{INR}_1} + \text{INR}_2 \right) \\
& + \log \left( 1 + \text{SNR}_1 + \text{INR}_1 + 2\sqrt{\text{SNR}_1 \cdot \text{INR}_1} \right) \\
& + p_2 \log \left( \frac{(1 + \text{INR}_2) \left(1 + \frac{\text{SNR}_2}{1 + \text{INR}_1}\right)}{1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} + \text{INR}_2} \right) \\
& - 2 \log 3 \\
& \geq (60) - 2 \log 3 - 1 - 2p_2
\end{aligned}$$

In order to bound  $\delta_2$ , we note that it cannot be larger than the maximum of the gaps found above, after normalization as done with bounding  $\delta_1$ . Hence, we find

$$\delta_2 < \log 3 + p_1 + p_2 \text{ bits.}$$

### C. Comparison with Suh-Tse Outer Bound

In this subsection, we compare our inner bound for the case  $p_1 = p_2 = 1$  with the perfect feedback outer bound of [1]. Looking at the region (11)–(14), we see that if we set  $p_1 = p_2 = 1$ , then the bounds (13) and (14) become redundant, and the region reduces to the outer bound region of [1], with the following differences:

- The outer bounds in [1] are parameterized by the parameter  $\rho$ , which captures the correlation between the symbols of two users. In the region (11)–(14), supremum values over all possible values of  $\rho$  is given.
- The bounds in (11)–(14) do not contain beamforming gain terms  $2\rho\sqrt{\text{SNR}_i \cdot \text{INR}_i}$ , which appear in the outer bounds of [1].

It is easy to see that the first item does not result in a rate penalty, while the second one gives a penalty of  $\log 3$  since

$$\begin{aligned}
(11)_L &= \log(1 + \text{SNR}_1 + \text{INR}_1) \\
&\geq \log \left( 1 + \text{SNR}_1 + \text{INR}_1 + 2\sqrt{\text{SNR}_1 \cdot \text{INR}_1} \right) - \log 3
\end{aligned}$$

Hence, the region  $\bar{\mathcal{C}}(1, 1)$  is within at most  $\log 3$  bits of the outer bound region of [1]. From the results of this section, we also know that our scheme achieves the region  $\bar{\mathcal{C}}(1, 1) - \delta_1$ . Evaluating  $\delta_1$  for  $p_1 = p_2 = 1$ , we see that the proposed

scheme achieves within  $3 + 3 \log 3 \approx 7.75$  bits of the Suh-Tse outer bound region.

## APPENDIX G

### PROOFS OF COROLLARIES 3.1 AND 3.3

In this section, we prove that when the feedback probabilities are sufficiently high, perfect feedback sum-capacity can be achieved (approximately for Gaussian case, exactly for linear deterministic case). The precise statements for the two models are given in Corollaries 3.1 and 3.3.

#### A. Proof of Corollary 3.1

We will show that when feedback is perfect, the bounds on the sum rate that involve the feedback probabilities become strictly redundant. That is, setting  $p_1 = p_2 = p$ , we will prove that the bounds (4), (1) + (2),  $\frac{(1)+(6)}{2}$ ,  $\frac{(2)+(5)}{2}$ , and  $\frac{(5)+(6)}{3}$  are all strictly larger than the perfect feedback bound (3) when  $p = 1$  and  $n_{12}, n_{21} > 0$ . Then the result follows by noting that all such bounds are continuous and monotonically increasing functions of  $p$ , and hence there must exist a  $p^* < 1$  such that whenever  $p = p^*$ , perfect feedback sum-rate bound (3) is exactly matched by these bounds.

We first prove a claim that will be used in the main proof.

**Claim G.1.** For  $n_{21}, n_{12} > 0$ ,

$$(3) < n_{12} + n_{21} + (n_{11} - n_{21})^+ + (n_{22} - n_{12})^+$$

*Proof.*

$$\begin{aligned}
& n_{12} + n_{21} + (n_{11} - n_{21})^+ + (n_{22} - n_{12})^+ \\
&= n_{12} + \max(n_{11}, n_{21}) + (n_{22} - n_{12})^+ \\
&= \max(n_{12} + n_{11}, n_{12} + n_{21}) + (n_{22} - n_{12})^+ \\
&> \max(n_{11}, n_{12}) + (n_{22} - n_{12})^+ \geq (3)
\end{aligned}$$

where the strict inequality follows by the fact that  $n_{21}, n_{12} > 0$ .  $\square$

Next, we consider the bound (4).

$$\begin{aligned}
(4) &= \max \left\{ n_{12}, (n_{11} - n_{21})^+ \right\} + \max \left\{ n_{21}, (n_{22} - n_{12})^+ \right\} \\
&+ \min \left\{ n_{12}, (n_{11} - n_{21})^+ \right\} + \min \left\{ n_{21}, (n_{22} - n_{12})^+ \right\} \\
&= n_{12} + n_{21} + (n_{11} - n_{21})^+ + (n_{22} - n_{12})^+ \\
&> (3)
\end{aligned}$$

where the last line follows by Claim G.1. Hence, the bound (4) becomes strictly redundant. Next, consider

$$\begin{aligned}
(5) + (6) &= \max(n_{11}, n_{12}) + \max(n_{22}, n_{21}) \\
&+ (n_{11} - n_{21})^+ + (n_{22} - n_{12})^+ \\
&+ \max \left\{ n_{12}, (n_{11} - n_{21})^+ \right\} \\
&+ \max \left\{ n_{21}, (n_{22} - n_{12})^+ \right\} \\
&+ \min \left\{ n_{12}, (n_{11} - n_{21})^+ \right\}
\end{aligned}$$

$$\begin{aligned}
& + \min \left\{ n_{21}, (n_{22} - n_{12})^+ \right\} \\
& = \max(n_{11}, n_{12}) + \max(n_{22}, n_{21}) \\
& \quad + 2(n_{11} - n_{21})^+ + 2(n_{22} - n_{12})^+ n_{12} + n_{21} \\
& \geq 2 \min \left\{ \max(n_{11}, n_{12}) + (n_{22} - n_{12})^+, \right. \\
& \quad \left. + \max(n_{22}, n_{21}) + (n_{11} - n_{21})^+ \right\} \\
& \quad + (n_{11} - n_{21})^+ + (n_{22} - n_{12})^+ n_{12} + n_{21} \\
& > 3 \cdot (3)
\end{aligned}$$

where the last line follows by Claim G.1.

Next, we consider the sum of individual rate bounds. Since these bounds consist of the minimum of two terms, we consider each case separately. In what follows,  $(1)_R$  denotes the term on the right-hand side of the minimization in (1), while  $(1)_L$  denotes the term on the left-hand side ( $(2)_R$  and  $(2)_L$  are also defined similarly). By symmetry, it is sufficient to prove that  $(1)_R + (2)_R$  and  $(1)_L + (2)_R$  are strictly redundant for  $p_1 = p_2 = 1$ . We show this as follows.

$$\begin{aligned}
(1)_R + (2)_R & = n_{11} + n_{22} + (n_{21} - n_{11})^+ \\
& \quad + (n_{12} - n_{22})^+ \\
& = \max(n_{11}, n_{21}) + \max(n_{22}, n_{12}) \\
& = n_{12} + n_{21} + (n_{11} - n_{21})^+ + (n_{22} - n_{12})^+ \\
& > (3)
\end{aligned}$$

by Claim G.1, and

$$\begin{aligned}
(1)_L + (2)_R & = \max(n_{11}, n_{12}) + n_{22} + (n_{12} - n_{22})^+ \\
& = \max(n_{11}, n_{12}) + \max(n_{22}, n_{12}) \\
& > \max(n_{11}, n_{12}) + (n_{22} - n_{12})^+ \geq (3)
\end{aligned}$$

since  $n_{12} > 0$ .

Finally, we consider the bounds (5) + (2), and (6) + (1). By symmetry, it is sufficient to show the redundancy of (5) + (2)<sub>R</sub> and (5) + (2)<sub>L</sub>. The former is shown by

$$\begin{aligned}
(5) + (2)_R & = \max(n_{11}, n_{12}) + (n_{11} - n_{21})^+ \\
& \quad + \max \left\{ n_{21}, (n_{22} - n_{12})^+ \right\} \\
& \quad + \min \left\{ n_{21}, (n_{22} - n_{12})^+ \right\} + n_{22} \\
& \quad + (n_{12} - n_{22})^+ \\
& = \max(n_{11}, n_{12}) + (n_{11} - n_{21})^+ \\
& \quad + n_{21} + (n_{22} - n_{12})^+ + \max(n_{22}, n_{12}) \\
& = \max(n_{11}, n_{12}) + (n_{22} - n_{12})^+ \\
& \quad + n_{21} + n_{12} + (n_{11} - n_{21})^+ (n_{22} - n_{12})^+ \\
& \geq (3) + n_{21} + n_{12} + (n_{11} - n_{21})^+ (n_{22} - n_{12})^+ \\
& > 2 \cdot (3)
\end{aligned}$$

where the last line follows by Claim G.1, and

$$\begin{aligned}
(5) + (2)_L & = \max(n_{11}, n_{12}) + (n_{11} - n_{21})^+ \\
& \quad + \max \left\{ n_{21}, (n_{22} - n_{12})^+ \right\} + \min \left\{ n_{21}, (n_{22} - n_{12})^+ \right\} \\
& \quad + \max \{ n_{22}, n_{21} \}
\end{aligned}$$

$$\begin{aligned}
& = \max(n_{11}, n_{12}) + (n_{11} - n_{21})^+ \\
& \quad + n_{21} + (n_{22} - n_{12})^+ + \max(n_{22}, n_{21}) \\
& > \max(n_{11}, n_{12}) + (n_{11} - n_{21})^+ \\
& \quad + \max(n_{22}, n_{21}) + (n_{22} - n_{12})^+ \\
& \geq 2 \min \left\{ \max(n_{11}, n_{12}) + (n_{22} - n_{12})^+, \right. \\
& \quad \left. + \max(n_{22}, n_{21}) + (n_{11} - n_{21})^+ \right\} \\
& = 2 \cdot (3)
\end{aligned}$$

where the strict inequality follows by the fact that  $n_{21} > 0$ .

### B. Proof of Corollary 3.3

Similar to the proof of Corollary 3.1, we will show that when feedback is perfect, the bounds on the sum rate that involve the feedback probabilities become redundant for the set  $\bar{\mathcal{C}}(p_1, p_2)$ . Since the capacity region is within a constant gap of the region  $\bar{\mathcal{C}}(p_1, p_2)$ , for all channel parameters, the result will follow.

Specifically, setting  $p_1 = p_2 = p$ , we will show that when  $p = 1$ ,  $\text{INR}_1, \text{INR}_2 > 0$ , the bounds (13),  $(11)_{(1,2)} + (11)_{(2,1)}$ ,  $\frac{(11)_{(1,2)} + (14)_{(2,1)}}{2}$ ,  $\frac{(11)_{(2,1)} + (14)_{(1,2)}}{2}$ , and  $\frac{(14)_{(1,2)} + (14)_{(2,1)}}{3}$  are all strictly larger than the perfect feedback bound (12), where the subscript  $(a, b)$  denotes the evaluation of the relevant bound with  $(i, j) = (a, b)$ .

We first prove a claim that will be useful in the proof of the corollary.

**Claim G.2.** For  $\text{INR}_1, \text{INR}_2 > 0$ ,

$$\begin{aligned}
\min_{(i,j)=(1,2),(2,1)} (12) & < \log(1 + \text{INR}_1) + \log(1 + \text{INR}_2) \\
& + \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) + \log \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right)
\end{aligned}$$

$$\begin{aligned}
& \log(1 + \text{INR}_1) + \log(1 + \text{INR}_2) \\
& + \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) + \log \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right) \\
& = \log \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right) \\
& + \log(1 + \text{INR}_1 + \text{INR}_2 + \text{SNR}_1 + \text{INR}_1 \text{INR}_2 + \text{INR}_1 \text{SNR}_1) \\
& > \log(1 + \text{SNR}_1 + \text{INR}_1) + \log \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right) \\
& \geq \min_{(i,j)=(1,2),(2,1)} (12)
\end{aligned}$$

Next, we show that under the condition  $\text{INR}_1, \text{INR}_2 > 0$  and  $p = 1$ , all of the mentioned bounds are strictly redundant. We start with (13):

$$\begin{aligned}
(13) & = \log(1 + \text{INR}_1) + \log(1 + \text{INR}_2) \\
& \quad + \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) + \log \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right) \\
& > \min_{(i,j)=(1,2),(2,1)} (12)
\end{aligned}$$

by Claim G.2. Next,

$$(14)_{(1,2)} + (14)_{(2,1)} = \log(1 + \text{SNR}_1 + \text{INR}_1)$$

$$\begin{aligned}
& + \log(1 + \text{SNR}_2 + \text{INR}_2) + 2 \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
& + 2 \log \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right) + \log(1 + \text{INR}_1) \\
& + \log(1 + \text{INR}_2) \\
& = (12)_{(1,2)} + (12)_{(2,1)} + \log(1 + \text{INR}_i) + \log(1 + \text{INR}_j) \\
& + \log \left( 1 + \frac{\text{SNR}_i}{1 + \text{INR}_j} \right) + \log \left( 1 + \frac{\text{SNR}_j}{1 + \text{INR}_i} \right) \\
& > 2 \cdot (12)_{(1,2)} + (12)_{(2,1)} \geq 3 \cdot \min_{(i,j)=(1,2),(2,1)} (12)
\end{aligned}$$

by Claim G.2. Next, we consider the bounds  $(11)_{(1,2),R} + (11)_{(2,1),R}$ ,  $(11)_{(1,2),L} + (11)_{(2,1),R}$ , and  $(11)_{(1,2),R} + (11)_{(2,1),L}$ , where the subscript  $L$  and  $R$  refer to the left-hand and right-hand side of the minimization in the relevant bounds, respectively. By symmetry, it is sufficient to show the redundancy of the former two.

$$\begin{aligned}
(11)_{(1,2),R} + (11)_{(2,1),R} &= \log(1 + \text{INR}_1) + \log(1 + \text{INR}_2) \\
& + \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) + \log \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right) \\
& > \min_{(i,j)=(1,2),(2,1)} (12)
\end{aligned}$$

by Claim G.2, and

$$\begin{aligned}
(11)_{(1,2),L} + (11)_{(2,1),R} &= \log(1 + \text{SNR}_1 + \text{INR}_1) \\
& + \log(1 + \text{SNR}_2 + \text{INR}_1) \\
& > \log(1 + \text{SNR}_1 + \text{INR}_1) + \log \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right) \\
& \geq \min_{(i,j)=(1,2),(2,1)} (12)
\end{aligned}$$

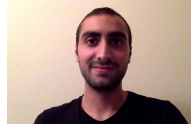
since  $\text{INR}_1, \text{INR}_2 > 0$ . Finally, we show the redundancy of  $\frac{(11)_{(1,2)} + (14)_{(2,1)}}{2}$  and  $\frac{(11)_{(2,1)} + (14)_{(1,2)}}{2}$ . By symmetry, we only consider the former, and in particular,  $\frac{(11)_{(1,2),L} + (14)_{(2,1)}}{2}$  and  $\frac{(11)_{(1,2),R} + (14)_{(2,1)}}{2}$ , where the subscript  $L$  and  $R$  refer to the left-hand and right-hand side of the minimization in the relevant bounds, respectively. Then

$$\begin{aligned}
(11)_{(1,2),L} + (14)_{(2,1)} &= \log(1 + \text{SNR}_1 + \text{INR}_1) \\
& + \log(1 + \text{SNR}_2 + \text{INR}_2) + \log \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right) \\
& + \log(1 + \text{INR}_1) + \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
& > \log(1 + \text{SNR}_1 + \text{INR}_1) \\
& + \log(1 + \text{SNR}_2 + \text{INR}_2) + \log \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right) \\
& + \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) = (12)_{(1,2)} + (12)_{(2,1)} \\
& \geq \min_{(i,j)=(1,2),(2,1)} (12)
\end{aligned}$$

and

$$\begin{aligned}
(11)_{(1,2),R} + (14)_{(2,1)} &= \log(1 + \text{INR}_2) \\
& + \log(1 + \text{SNR}_2 + \text{INR}_2) + \log \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right) \\
& + \log(1 + \text{INR}_1) + 2 \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right)
\end{aligned}$$

$$\begin{aligned}
& > (12)_{(2,1)} + \log(1 + \text{INR}_2) + \log(1 + \text{INR}_1) \\
& \quad \log \left( 1 + \frac{\text{SNR}_2}{1 + \text{INR}_1} \right) + \log \left( 1 + \frac{\text{SNR}_1}{1 + \text{INR}_2} \right) \\
& > 2 \cdot \min_{(i,j)=(1,2),(2,1)} (12).
\end{aligned}$$



**Can Karakus** received his B.S. degree from Bilkent University, Turkey, in 2011, his M.S. degree from University of California, Los Angeles (UCLA), USA, in 2013, and currently working towards his Ph.D. degree at UCLA, all in electrical engineering. He was a recipient of UCLA Graduate Division Fellowship in 2011, and UCLA Electrical Engineering Department Fellowship in 2013. His research interests include information theory, wireless networks and algorithms.



**I-Hsiang Wang** received the B.Sc. degree in electrical engineering from National Taiwan University, Taiwan, in 2006. He received a Ph.D. degree in electrical engineering and computer sciences from the University of California at Berkeley, USA, in 2011. From 2011 to 2013, he was a postdoctoral researcher at École Polytechnique Fédérale de Lausanne, Switzerland. Since 2013, he has been at the Department of Electrical Engineering in National Taiwan University, where he is now an assistant professor. His research interests include network information theory, wireless networks, coding theory, and network coding. He received a 2-year Vodafone Graduate Fellowship in 2006. He was a finalist of the Best Student Paper Award of IEEE International Symposium on Information Theory, 2011.



**Suhas Diggavi** received the B. Tech. degree in electrical engineering from the Indian Institute of Technology, Delhi, India, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1998.

After completing his Ph.D., he was a Principal Member Technical Staff in the Information Sciences Center, AT&T Shannon Laboratories, Florham Park, NJ. After that he was on the faculty of the School of Computer and Communication Sciences, EPFL, where he directed the Laboratory for Information and Communication Systems (LICOS). He is currently a Professor, in the Department of Electrical Engineering, at the University of California, Los Angeles, where he directs the Information Theory and Systems laboratory. His research interests include wireless network information theory, wireless networking systems, network data compression and network algorithms.

He is a co-recipient of the 2013 IEEE Information Theory Society & Communications Society Joint Paper Award, the 2013 ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc) best paper award, the 2006 IEEE Donald Fink prize paper award, 2005 IEEE Vehicular Technology Conference best paper award and the Okawa foundation research. He is a Fellow of the IEEE, a distinguished lecturer for the information theory society and has served on the editorial board for Transactions on Information Theory, ACM/IEEE Transactions on Networking, IEEE Communication Letters, a guest editor for IEEE Selected Topics in Signal Processing. He served as the Technical Program Co-Chair for 2012 IEEE Information Theory Workshop (ITW) and the Technical Program Co-Chair for the 2015 IEEE International Symposium on Information Theory (ISIT).